THE UNIVERSITY *of* EDINBURGH

A Statistical Method for Identification of Sources of Electromechanical Oscillations in Power Systems

Patrick McNabb

Thesis appendices submitted for the degree of Doctor of Philosophy

College of Science and Engineering

University of Edinburgh

February 2011

Abstract

The use of real-time continuous dynamics monitoring often indicates dynamic behaviour that was not anticipated by model-based studies. In such cases it can be difficult to locate the sources of problems using conventional tools. This thesis details the possibility of diagnosing the causes of problems related to oscillatory stability using measurement-based data such as active power and mode decay time constant, derived from system models. The aim of this work was to identify dynamics problems independently of an analytical dynamic model, which should prove useful in diagnosing and correcting dynamics problems.

New statistical techniques were applied to both dynamic models and real systems which yielded information about the causes of the long decay time constants observed in these systems. Wavelet transforms in conjunction with General Linear Models (GLMs) were used to improve the statistical prediction of decay time constants derived from the system. Logic regression was introduced as a method of establishing important interactions of loadflow variables that contribute to poor damping.

The methodology was used in a number of case studies including the 0.62Hz Icelandic model mode and a 0.48Hz mode from the real Australian system. The results presented herein confirm the feasibility of this approach to the oscillation source location problem, as combinations of loadflow variables can be identified and used to control mode damping. These ranked combinations could be used by a system operator to provide more comprehensive control of oscillations in comparison to current techniques.

Declaration Of Authorship

I, Patrick McNabb, declare that the the work presented in this thesis is my own, and has been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signed: .....

Date: .....

Acknowledgments

I would like to thank my principal supervisor Janusz Bialek, who gave me the opportunity and support to carry out the research contained in this thesis. Without his technical and collaborative contributions this thesis would not have been possible.

I am eternally grateful to my industrial supervisor Douglas Wilson who took the time to sit down with me and explain some of the problems I encountered along the way. I hope I can repay him with a successful implementation of the work in this thesis.

My gratitude also goes to Zbigniew Lubosny who helped me in the early stages of my PhD, particularly with developing the models and solving the significant problems I had with them.

A special thank you goes to the people at Psymetrix who co-funded this work and who allowed me to use their PhasorPoint software for small-signal stability analysis, in addition to providing me with data from Australia and Iceland.

This work couldn't have been done without the invaluable help of Natalia Bochkina, a statistics lecturer at the University of Edinburgh, who helped me with the many statistical problems I faced during this work.

Finally, I would like to acknowledge the EPSRC who funded this work along with Psymetrix.

kgkgkg

lglglg

lglgl

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The phenomenon of oscillatory instability has been recognized in the power industry for a long time and there are proven techniques to control system oscillations. Of major concern today are the inter-area oscillations that severely constrict the transfer of power between areas to meet load demands. In principle, control techniques that provide damping forces can be applied to various types of plant, most notably at large central generators. In practice, it is difficult to design and co-ordinate such controls so that they work over all practical operating scenarios. Since practical modeling is at best an approximation to the real-world behaviour, there can be unexpected behaviour in the real system's response to disturbances [1].

To achieve the goal of maintaining stability, the system must be operated with a suitable margin of safety, so unexpected or hidden problems do not result in major system disturbances. A careful judgment must be made in order to operate the system securely, but without incurring undue costs by over-constraining the network. In order to achieve this balance, it is important for network owners and operators to have accurate and timely information on the security of the system.

There are a number of further challenges to the task of maintaining oscillatory stability that have arisen over the recent years, including:

- Power markets largely determine the generation that operates at a given time. Since the generators provide a major contribution to dynamic performance, the margin of stability is significantly affected by market conditions. A lively power market will push the system to many operating states, some of which were not considered in system planning

- Commercial sensitivity created by separating generation from transmission means that opportunities to carry out dynamic testing and tuning of control systems are restricted

- Environmental and commercial pressures tend to push the power system closer to its physical limits of power transfers. This contributes to stress on the system that can expose problems that would otherwise remain hidden

- Different generation technologies have different dynamic responses. At present, the control of oscillations is largely carried out using large conventional generators. However, as renewable and distributed generation are responsible for an increasing share of generation, the dynamic characteristics of the power system will change

- Expansion of synchronously connected regions significantly changes the dynamics of the network. For example, the recent European network (UCTE) extension into the Balkans and North Africa increases the complexity of interactions over a wide area

- There is pressure on the analysis and planning of resources available to TSO's

Recent developments in measurement and processing techniques have been used to monitor the stability of oscillations directly. Combined with fast, continuous real-time data monitoring, this technology enables real-time monitoring of the margin of safety maintained against potential instability. The technology has been used to detect and warn of the system approaching instability [2].

The use of real-time continuous dynamics monitoring often indicates dynamic behaviour that was not reproduced in model-based studies. In such cases, it can be

difficult to track down problems using conventional tools. This thesis explores the possibility of diagnosing problems related to oscillatory stability using measurement-based techniques. Methods for identifying dynamics problems independently of an analytical dynamic model would be very useful in diagnosing and correcting a range of dynamics problems.

In order to provide a suitable solution to the source location problem, a database of system states including active and reactive power flows, bus voltages, active and reactive load and generation as well as the system eigenvalues are required to estimate significant predictors and combinations thereof. This process can then be used to determine the locations of likely sources of oscillations.

In order to maintain the system stability, the system must be operated with a suitable margin of safety, so unexpected or hidden problems do not result in major disturbances. A careful judgment must be made in order to operate the system securely, but without incurring undue costs by over-constraining the network. In order to achieve this balance, it is important for network owners and operators to have accurate and timely information on the security of the system.

1.2 Research Objectives and Scope

The objectives of this research are as follows:

- Describe and illustrate the various common mechanisms of oscillatory behaviour in power systems, and describe the observable features of each mechanism
- Develop suitable signal processing techniques for use in subsequent procedures
- Develop and illustrate statistical procedures to be used in the diagnosis of oscillation problems occurring in an interconnected network
- Using case studies, demonstrate the application of the diagnosis methods to observed power system studies

- Draw conclusions on the practical application of observation and diagnosis techniques, outlining measurement requirements
- Develop models to test any methodology for use in a measurement-based system

1.3 Thesis Statement and Contribution to Knowledge

Using measurements from wide-area measurement systems and SCADA, it should be possible using a number statistical techniques to identify correlations between loadflow data and mode damping data that can be used to identify sources of electromechanical oscillations in power systems. The contributions to knowledge from this research are as follows:

- To design a methodology for the location of the source of electromechanical oscillations
- To derive signal processing techniques, such as the wavelet transform, for use in subsequent statistical procedures as well as for event detection
- Use Psymetrix PhasorPoint application to monitor the system i.e. continually extract decay time constant data from real active power signals measured via PMU to determine if it exceeds a set threshold.
- Use decay time constant data SCADA data to create statistical models for relating the combinational effects of predictors to the response
- Collaborate with a statistician to develop general linear and logic regression models to relate predictors to response variables in the models
- To provide a real-time indication of source location for utilities and TSO's so that they can respond immediately to unstable oscillations
- As a result of the work detailed above, TSO's will be able to maintain real-time dynamic stability and secure operation of continuous electric power supply in response to electromechanical oscillations. With the use of PhasorPoint and this source location technique, system operators will be able to access information on

the causes of poorly damped oscillations and will be in a position to take remedial action

1.4 Thesis Outline

In chapter 2, the background of power systems and their stability problems are discussed with emphasis on some of the analysis and detection methods of troublesome dynamics. The interpretation of important mathematical concepts such as eigenvectors and participation factors are covered as well which are used in the event detection process and in the source location methodology.

Chapter 3 is a literature survey describing some major events worldwide as well as the solutions and courses of action taken. Other methodologies are presented as possible solutions to the source location problem which vary from neural networking solutions to purely statistical approaches. Chapter 4 is a report on the models used and modeling techniques employed to replicate some of the behaviour in the real system for use in testing.

Chapter 5 introduces signal processing via the wavelet transform as a means to regionalize the signal and prepare the data for subsequent statistical steps. Generalized linear models are then discussed as a method introducing non-linearity into the statistical models. General linear models (GLM's), which are a type of generalized linear model, are then combined with wavelet transformed data, in an attempt to produce better fitted statistical models with more accurate predictions.

Chapter 6 applies the logic regression methodology to a number of power system models for validation. This is applied to both raw and wavelet transformed data to determine the best model fits. The methodologies presented in chapter 5 and 6 are then applied to a model case study as well as real system data in chapter 7. Results are presented that show the usefulness of both raw and wavelet-transformed approaches. Both are validated through model testing as well as through the comparison of system states

and corresponding dynamics.

Decay time constant and decay time both refer to the quantity expressed in Equation 2.2, i.e. the negative of the reciporical of the real part of a mode eigenvalue. Physically, this is the time it takes for an mode to decay to $1/\ln(1) = 0.367$ of its initial amplitude.

1.5 System Outline

Figure 1.1 shows a block diagram of the source location algorithm that will be described and explained in the remainder of this thesis. This is a useful diagram to reference when reading through the various functions in the methodology as it can be difficult to piece together the components to visualize the final system.

The diagram begins with inputs, consisting of system state data (active/reactive power etc.) and dynamics data (decay time constant/damping ratio). This data is derived from both SCADA and phasor measurement units (PMU's) via WAMS. PMU's measure current and voltage phasors at various locations in the network at 50Hz. These high resolution measurements are processed using Psymetrix's phasorpoint software to derive mode attributes such as decay time constant and damping ratio. The SCADA system measures much lower resolution values of current and voltage but typically at a larger number of locations than the PMU's.
This data is thresholded and wavelet transformed and the resulting raw and wavelet transformed data is again thresholded using mutual information and partial least squares regression.

After optimal scales are selected from the wavelet transformed data, it is then used to develop a general linear model with the output consisting of a set of system variables ranked in terms of their effect on damping.
The wavelet data along with the raw data is binary thresholded to create binary variables from the original data for use in logic regression. By using this binary data and the output from the general linear model, a ranked list of combinations of system vari-

ables (ranked in terms of effect on damping) can be used to control the damping of certain electromechanical modes.

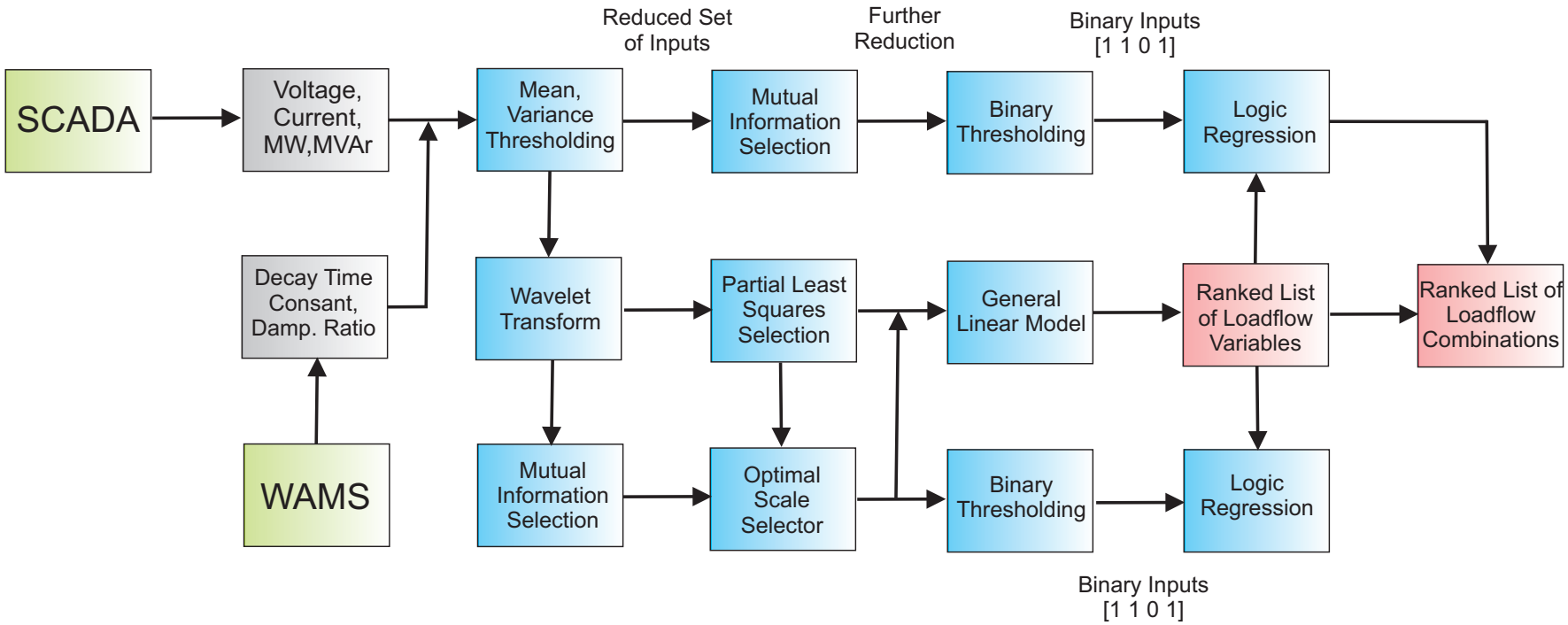


Figure 1.1: Block diagram of Source Location Algorithm. High resolution current and voltage inputs are measured via Phasor Measurement Units (PMU's) that are part of Wide Area Measurement Systems (WAMS). SCADA measurements supply lower resolution loadflow variables at more locations in the system

Chapter 2

Background

2.1 Introduction

Todays increasing demand for maximum power transfer in many parts of the world is growing faster than power utilities can handle. As a result of deregulation, independent power utilities are connected together over extremely long transmission lines that are transferring large amounts of bulk power to different regions to meet local energy demands. In addition, such large interconnections of different control regions can provide a means to enhance network security and economy of operation [3]. However, these interconnections come at a cost to the power system as more complex networks have more complex dynamics associated with them.

Unfortunately, the complexity of the stability problems associated with maintaining optimal power system security and reliable operation has amplified as well. In this chapter, low frequency oscillations and some power system history is presented as well as some of the existing solutions dealing with stability problems experienced in todays power systems.

2.2 Oscillations in Power Systems

Maintaining power system stability is a complex field that takes into account many different physical phenomena. It is also an important responsibility, as failure to maintain stability can have very serious consequences for the utilities concerned as well as

their customers.

A joint CIGRE/IEEE taskforce recently defined and classified power system stability broadly along the lines as shown in figure 2.1.

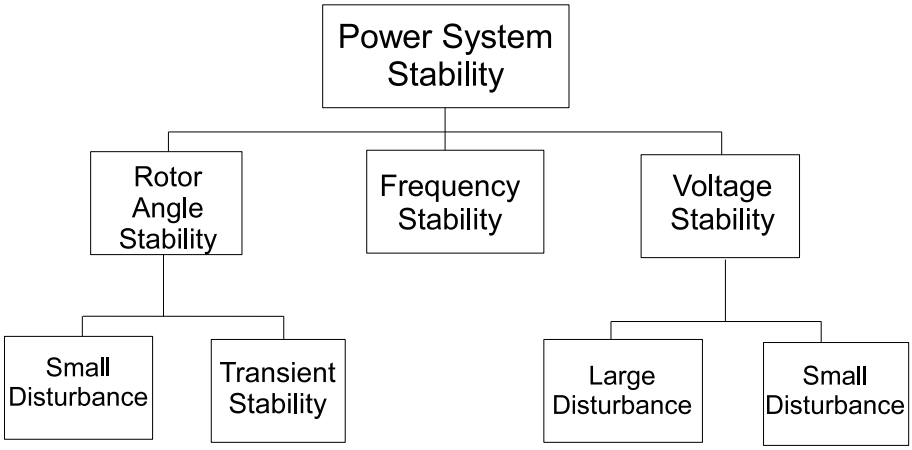


Figure 2.1: Classification of Power System Stability

Oscillatory stability is part of rotor angle stability, and is concerned with whether there is sufficient damping in a network that would cause oscillations triggered by a disturbance to die away. While the damping of oscillations is normally associated with small-disturbance angle stability, the response to transients is also affected by damping.

Of particular concern are electromechanical oscillations, which are natural dynamic resonances between generators interconnected by a transmission system, analogous to inertial masses connected by springs. All power systems experience electromechanical oscillations. These resonances can be local in nature, involving only one rotor or a localized group of generators, or they can involve a large number of generators over a wide area as is the case in interarea mode oscillations.

Typically, the oscillations occur at frequencies between 0.2Hz and 1.5Hz, with inter-area modes at lower frequencies and local modes at higher frequencies. The key to managing oscillatory stability is to make sure that all oscillatory modes in the system are adequately damped at all times.

2.3 Analyzing Oscillatory Stability

There are two complementary approaches to analyzing oscillatory stability. The more conventional approach is to use dynamic modeling to quantify damping and to design measures to solve stability problems. In particular, small-disturbance stability methods continue to be important tools. Recently, measurement-based techniques have been developed and used effectively by utilities to monitor the stability of the network and ensure that a suitable safety margin is maintained against oscillatory instability.

Model-based approaches are useful for understanding the underlying conditions and designing damping controllers. However, it is shown in the literature review that they are not always useful for predicting instability. Measurement-based techniques are useful for detecting when stability problems arise and warning operators before they become critical. However, measurement-based techniques are not predictive and it can be difficult to determine the root causes of oscillations using only these techniques. Industry best practice involves the use of both model and measurement-based methodologies.

The ultimate goal of this work is the development of a methodology to solve oscillatory stability problems on the basis of real measurements. In real world observations however, the true underlying system conditions are not known precisely. Many of the illustrations in this thesis are based on simulations of stability phenomena using small-disturbance modeling techniques. The features of the various conditions are then examined in observable quantities. The basis for linear modeling and measurement-based analysis are considered in the following subsections.

2.4 Small-Disturbance Modeling of Oscillations

The goals of small-disturbance modeling are broadly:

- To determine if there are any unstable modes of oscillation in a system, or modes that may become unstable after a credible contingency
- To determine the nature of oscillations in the network and identify where the

oscillations are observable and controllable and which generators participate most strongly

- To design control systems in such a way that oscillations are well damped

Small-disturbance modeling techniques are well established and described in several texts [4] [5]. The following description is intended only as an outline to summarize the steps involved.

2.5 Representation of System Equations in State-Space Form

The behaviour of a power system can be described by a set of n first-order non-linear ordinary differential equations and m non-linear algebraic equations, expressed in vector format as follows:

$$\begin{aligned} \dot{\underline{x}} &= f(\underline{x}, \underline{u}, t) \qquad differential\ equations \\ \underline{y} &= g(\underline{x}, \underline{u}, t) \qquad algebraic\ equations \end{aligned} \tag{2.1}$$

where:

$$\begin{aligned} \underline{x}: &\quad state\ variables \\ \underline{u}: &\quad input\ variables \end{aligned} \tag{2.2}$$

Examples of System States are:

- Generator rotor speed
- Generator rotor angle
- Flux in d-axis or q-axis

Examples of Inputs are:

- AVR reference voltage
- Turbine governor reference speed
- Loads

2.6 Linearization of State Equations around and Equilibrium Point

The non-linear equations below are linearized around the equilibrium points, at which $\underline{\dot{x}}$=0. At an equilibrium point, the state equations can be expressed as:

$$\Delta\underline{\dot{x}} = A\Delta\underline{x} + B\Delta\underline{u} \tag{2.3}$$

$$\Delta\underline{y} = C\Delta\underline{x} + D\Delta\underline{u} \tag{2.4}$$

Using a Taylor series expansion, neglecting higher order terms:

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} \cdots \frac{\partial f_1}{\partial x_n} \\ \vdots \ddots \vdots \\ \frac{\partial f_n}{\partial x_1} \cdots \frac{\partial f_n}{\partial x_n} \end{bmatrix} \qquad B = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} \cdots \frac{\partial f_1}{\partial u_n} \\ \vdots \ddots \vdots \\ \frac{\partial f_n}{\partial u_1} \cdots \frac{\partial f_n}{\partial u_n} \end{bmatrix} \tag{2.5}$$

$$C = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} \cdots \frac{\partial g_1}{\partial x_n} \\ \vdots \ddots \vdots \\ \frac{\partial g_m}{\partial x_1} \cdots \frac{\partial g_m}{\partial x_n} \end{bmatrix} \qquad D = \begin{bmatrix} \frac{\partial g_1}{\partial u_1} \cdots \frac{\partial g_1}{\partial u_r} \\ \vdots \ddots \vdots \\ \frac{\partial g_m}{\partial u_1} \cdots \frac{\partial g_m}{\partial u_r} \end{bmatrix} \tag{2.6}$$

Since the original state equations are non-linear, it may be noted that the linear A, B, C, D matrices apply to only one particular state of the system, and must be recalculated for each system state to be studied. For this reason, eigenvalue analysis can be computationally intensive, and techniques to reduce the computational effort are required in order to study large systems.

2.7 Determining Eigenvalues and Eigenvectors

The right eigenvectors associated with the eigenvalues are orthogonal vectors that transform the state model, such that $\underline{\dot{x_i}}$ becomes a scalar multiple of $\underline{x_i}$. In this special case:

$$Av_i = \lambda_i v_i \tag{2.7}$$

Eigenvalues $\lambda_1 \ldots \lambda_n$ are solutions to the equation:

$$det[A - \lambda_i I] = 0 \qquad for\ \ each\ \ i = 1 \ldots n \tag{2.8}$$

$$Av_i = \lambda_i v_i \qquad v_i : right\ \ eigenvector\ \ for\ \ ith\ \ mode \tag{2.9}$$

$$w_i A = w_i \lambda_i \qquad w_i : left\ \ eigenvector\ \ for\ \ ith\ \ mode \tag{2.10}$$

Only left and right eigenvectors for the same mode have a non-zero product, ie:

$$w_j v_i = 0 \qquad if \quad i \neq j \tag{2.11}$$

$$w_i v_i \neq 0 \qquad w_i v_i = 1 \quad after\ \ normalization \tag{2.12}$$

Thus, the normalized form can be expressed in matrix form:

$$AV = \Lambda V \tag{2.13}$$

$$V^{-1} AV = \Lambda \quad expressing\ \ W = V^{-1}\ \ gives\ \ WAV = \Lambda \quad WV = I \tag{2.14}$$

where:

$$V = \begin{bmatrix} \vdots \cdots \vdots \\ \underline{v_1} \underline{v_2} \cdots \underline{v_n} \\ \vdots \cdots \vdots \end{bmatrix} \tag{2.15}$$

$$W = \begin{bmatrix} \cdots \underline{w_1} \cdots \\ \cdots \underline{w_2} \cdots \\ \vdots \\ \cdots \underline{w_n} \cdots \end{bmatrix} \tag{2.16}$$

and:

$$\Lambda = \begin{bmatrix} \lambda_1 0 \cdots 0 \\ 0 \lambda_2 \cdots 0 \\ \vdots \ddots \vdots \\ 0 0 \cdots \lambda_n \end{bmatrix} \tag{2.17}$$

The time response of the ith state variable is given by:

$$\Delta x_i(t) = v_{il}\underline{w_1}\Delta\underline{x_0}e^{\lambda_1 t} + v_{i2}\underline{w_2}\Delta\underline{x_0}e^{\lambda_2 t} + \ldots + v_{in}\underline{w_n}\Delta\underline{x_0}e^{\lambda_n t} \tag{2.18}$$

where:

$$\lambda_i = \sigma_i \pm j\omega_i \tag{2.19}$$

and the initial conditions of the state variables are given by the vector $\Delta\underline{x_0}$. The $e^{\lambda_i t}$ term therefore comprises a damped exponential term ($e^{\sigma_i t}$) and an oscillatory term $e^{j\omega_i t}$, if the eigenvalue is complex. The frequency and damping of the mode are given by:

$$f = \frac{\omega}{2\pi} \tag{2.20}$$

$$\zeta = \frac{-\sigma}{\sqrt{\sigma^2 + \omega^2}} \tag{2.21}$$

Alternatively, damping can be expressed as:

$$\tau = \frac{-1}{\sigma} \qquad for\ \ small\ \ \zeta : \tau \approx \frac{1}{\omega\zeta} \tag{2.22}$$

It is also worth noting that:

- The observability of a mode defines the movement of the output variables in response to the mode. The ith mode is observable in the jth output if the product is non-zero. The relative magnitude of the products is known as the observability.

- The ith mode is controllable by the jth input if the product is non-zero. The relative magnitude of the products is known as the observability.

- The participation factors are a measure of the relative participation of the kth state variable in the ith mode, and vice versa. The participation factors measure the activity of the state variable x_k in the ith mode, weighted by the contribution of this activity to the mode. Participation factors for the ith mode are given by:

$$\underline{p_i} = \begin{bmatrix} v_{1i} & w_{i1} \\ v_{2i} & w_{i2} \\ \vdots & \vdots \\ v_{ni} & w_{in} \end{bmatrix} \tag{2.23}$$

2.8 Interpreting Eigenvalues and Eigenvectors

The eigenvalues can be real or complex, and the real part can be positive or negative. In terms of stability of the system, the eigenvalues can be interpreted as follows:

- If the real parts of all eigenvalues are negative, the system is small-disturbance stable at the operating point. The real part determines the damping, which is the rate at which the system returns to a stable operating point.
- An eigenvalue with a positive real part corresponds to an unstable mode; the system will not return to a stable equilibrium after a small disturbance
- An eigenvalue with a zero real part corresponds to an undamped mode; the system will continue to oscillate in response to a small disturbance
- The imaginary part corresponds to ω, a natural frequency or mode of oscillation of the system. If an eigenvalue is real, the response is non-oscillatory.

The eigenvalue analysis therefore provides a system analyst with a means to calculate the modes of oscillation in the system, and damping associated with each mode.
It may be noted that an eigenvalue with a small real part is of concern, as the response is close to instability. A relatively small change in the system could result in an unstable condition. System design criteria sometimes specify a margin of stability to account for modeling inaccuracy and potential system state changes.
Observability, controllability and participation factors can be interpreted as follows:

- The observability of a mode defines the relative amplitude and phase of amplitudes in the outputs in quantities that can be measured.

- The controllability of a mode by an input is the influence that an input has on the mode
- The participation factors are normally associated with generator rotor angles, and determine the relative contribution of the generators to a mode.

2.9 Measurement-Based Methods for Characterizing Oscillatory Behaviour

2.9.1 Damping Analysis of Large Disturbances

Conventionally, estimates of oscillation damping have been based on the system's response to a transient event. The transient event may be a fault, or a disturbance that is designed to excite the system.

It may be noted from the above discussion on the time response of the system that the expected response is the superposition of a number of damped sinusoidal components. Prony analysis is a technique to analyze ringdown responses, extracting the frequency and damping of dominant modes from the measured response [6]. Essentially, the method fits a series of exponentially damped sinusoids to a measured response, using a least squares fitting criterion.

There are, however, a number of problems with the use of Prony analysis in this way:

- Since the method assumes that the measured response is due to an initial disturbance, the presence of continuous perturbations of the system corrupts the results. A large perturbation is required so that the response to the transient is distinct. The presence of continuous exciting noise can produce an undamped or negatively damped result when the system is actually stable.
- Estimates of damping can only be obtained at specific times when there is a disturbance. The method is limited to snapshots of system behaviour.
- The estimation of damping assumes a linear system response. However, a relatively large disturbance is required for a distinct ringdown response, and the system may not respond to the disturbance as a linear second order system.

2.9.2 Damping Analysis of Small Disturbances

A detailed look at a measured power system signal shows a steady state component coupled with small disturbances from the steady state. A typical active power signal from a transmission line is shown in figure 2.2, with small oscillatory disturbances continuously being observed. These disturbances are the system's response to small random perturbations, mostly from small changes in loads around the grid. The characteristics of the small oscillatory disturbances in the measured signal contain information on the mode dynamics of the power system.

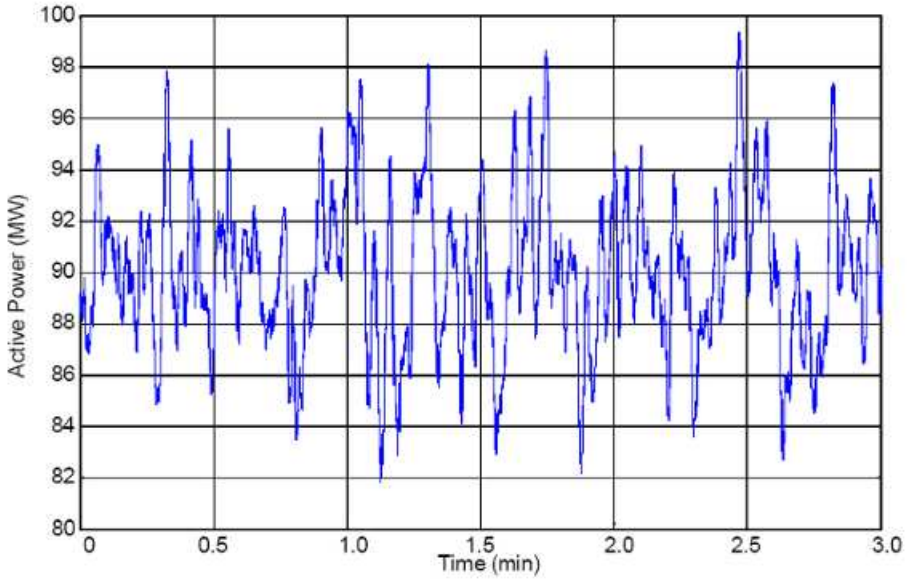


Figure 2.2: Section of Active Power signal with Small Disturbances

Approximating the power system as a second order linear dynamical system is accurate as long as the power system is subjected to small perturbations, rather than major changes in the operational state of the system.

For the system identification problem, the output of the system is one or more measured values such as active power flows in transmission lines, voltage angles or system frequency. Random changes at multiple load points are the inputs, as shown in figure 2.3. The proprietary analysis method developed by Psymetrix Ltd. derives the mode frequency, amplitude and damping of all dominant modes observable at the measured

output of the system.

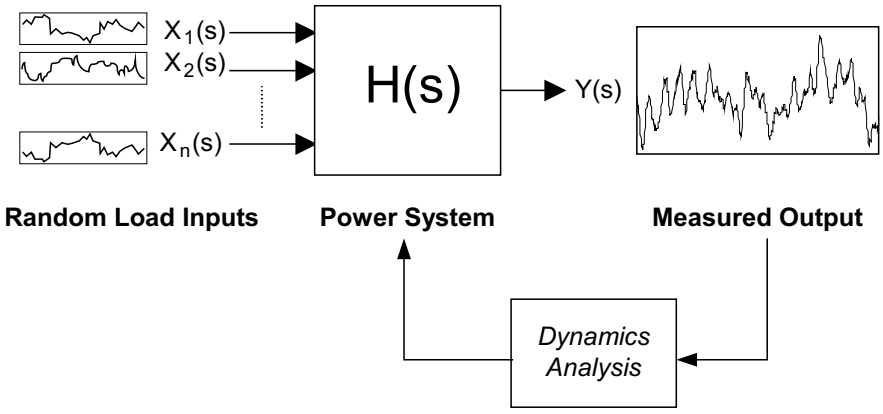


Figure 2.3: Transfer function representation of the power system with random inputs

2.10 Summary

Modeling techniques to deal with oscillatory stability are well established. However, there are still practical problems in dealing with oscillations in a power system in terms of accuracy of modeling and the choice of systems states for simulation. Unexpected problems can create either unstable or insecure conditions in the network. In the latter case, the problem may be undetected but still represent a weakness in the grid that may result in a serious disturbance given a particular event or system condition. Such a weakness was present in all of the case study examples which will follow in chapter 3, and the eventual result in these cases was severe stress on the power system, resulting in widespread blackouts in two of the cases.

In this chapter, methods to identify the frequency and damping from direct measurements were presented. To date, systems incorporating these methods have been mainly installed to deal with known problems where the underlying phenomenon is understood and suitable actions can be taken. However, the application of the dynamics monitoring technology often reveals other, less well-understood patterns of behaviour. In such cases, the reason for the reduced stability margin is seldom obvious, and it can be hard to identify an appropriate course of action.

There is a need to be able to characterize the behaviour of the system using measurements. This would help both system operators and analysts to determine an appropriate course of action. In the operational timeframe, this requires fast determination of an appropriate redispatch action. In the planning and analysis timeframe, the actions will typically involve model validation and damping controller design. In both cases, methods to characterize the dynamic behaviour and the sensitivities would greatly improve the understanding of the problem and the ability to take action to mitigate the risk. The purpose of this thesis is to investigate ways of using readily available data to identify the causes of oscillations, with a view to resolving the issues [2].

It may be concluded that there are considerable risks associated with oscillatory stability. The emergence of tools to manage oscillatory stability using a measurement-based approach is a very useful complement to the conventional model-based approaches. However, even where these tools detect stability risks, it can be very difficult to track down the causes of the problems or determine appropriate action. Methods explored in this work for recognizing the nature of problems and identifying suitable responses are therefore of practical importance.

Chapter 3

Literature Review

3.1 Introduction

Power system stability has been of major concern since the early days of interconnecting synchronous generators in different areas through high voltage transmission lines. Small signal instability in the early stages was related to the loss of synchronism due to lack of synchronizing torque. The development of fast, automatic voltage regulators (AVR's) acting on the synchronous generators excitation systems solved the problem of synchronism as the AVR increases synchronizing torque once it falls below a preset level.
However, remedying this low synchronizing torque leads to a deterioration in the damping of electromechanical oscillations of the weakly interconnected power system due to the negative damping characteristics of the AVR [4].

In order to damp troublesome modes of oscillation, decentralized power system stabilizers (PSSs) have been developed. The conventional power system stabilizers use local signals such as active power, rotor speed and/or frequency as input signals to provide a damping torque in phase with speed deviation to damp local and inter-area oscillations. The PSSs are time invariant controllers which are defined by transfer functions with constant parameters; the parameter of PSSs are determined using a linearized plant model (i.e. synchronous generator connected to the power system) around a given operating condition [7]. These distributed PSSs have proved to be useful in damping

local and interarea oscillations provided their parameters have been calculated for a wide range of operating conditions.

However, in recent years the major concern of power system utilities has shifted toward inter-area oscillations rather than local oscillations because of the great impact inter-area oscillations can have on the power system as a whole. The increase in inter-area oscillations can limit the amount of power exchange and can cause the break-up of the whole power system if they cannot be damped effectively. It is also possible that strong coupling may exist between local and inter-area oscillations which makes it extremely difficult to tune PSS's to be effective against these types of oscillations.

To achieve the goal of stability, the system must be operated with a suitable margin of stability, so unexpected or hidden problems do not result in major system disturbances. A careful judgment must be made in order to operate the system securely but without incurring undue costs by over-constraining the network. In order to achieve this balance, it is important for network owners and operators to have accurate and timely information on the security of the system.

Recent developments in measurement and processing techniques have been used to monitor the stability of oscillations directly. Combined with fast, continuous real-time data monitoring, this technology enables real-time monitoring of the margin of safety maintained against potential instability. The technology has been used to detect and warn of the system approaching instability.

The use of real-time continuous dynamics monitoring often indicates dynamic behaviour that was not expected from model based studies. In such cases, it can be difficult to track down problems using conventional tools. This thesis explores the possibility of diagnosing problems related to oscillatory stability using measurement based techniques. Methods for identifying dynamics problems independently of an analytical dynamic model would be very useful in diagnosing and correcting dynamics problems. This chapter illustrates a number of different problems associated with oscillatory sta-

bility and how these problems were dealt with using a number of methods to identify the causes of poor dynamics.

3.2 Examples of Observed Problems

Maintaining power system dynamic stability is a practical problem faced by utilities. In this section, a number of more severe incidents of oscillatory stability are presented. These events illustrate the nature of the problem and the extent of risk if this stability problem is not adequately managed. A historical perspective and further examples are provided in a CIGRE technical brochure [8].

3.2.1 United Kingdom, 1980

The UK grid is a dense network with relatively short lines. However, an interarea mode exists between coherent groups of generators is Scotland and England. On several occasions around 1980, the interarea mode became unstable [9]. The system became increasingly poorly damped as certain operating conditions were approached and instability occurred without changes in topology due to faults. The unstable condition led to 0.5Hz oscillations typically growing to 1000MW peak-to-peak observed on the interconnection between the areas. An example of the oscillations are shown in figure 3.1.

The unstable system could be restored to stability by reducing the level of the power flow on the interconnector. It could also be addressed by switching generator AVR controls to manual but this was a temporary response to an emergency rather than a long-term solution. The problem was addressed initially by implementing a transfer constraint, though unstable conditions still occurred.

The system did not separate on any of these occasions but severe wear on the generator shafts was noted. In addition to the financial cost associated with the transfer constraint, there was also a significant safety risk associated with the possibility of generator shaft breakage.

The condition was resolved by:

- Deploying stability monitoring at the interconnector. This included online damping monitoring, so that a stability margin could be maintained without an over-conservative constraint.
- Power system stabilizers configured to damp the 0.5Hz mode were deployed in selected Scottish generators.

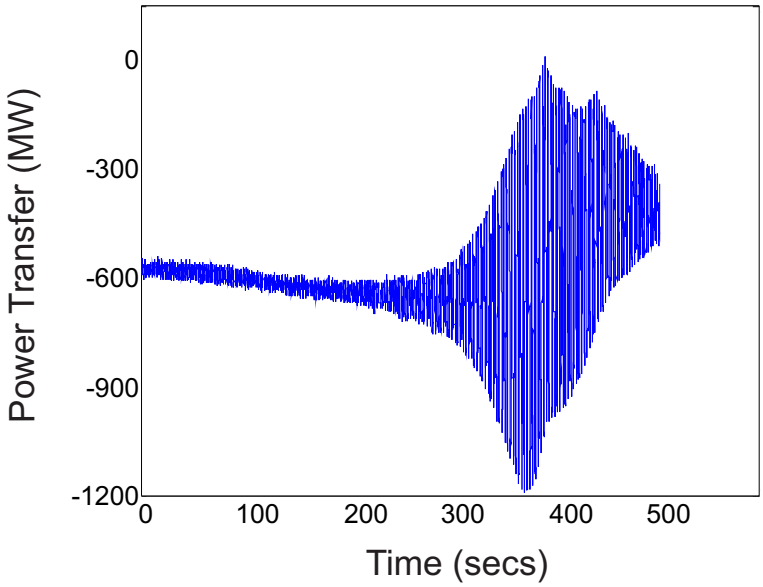


Figure 3.1: Unstable Power Oscillations, UK, 1980

3.2.2 West USA/Canada, System Separation, 1996

The blackout on August 10, 1996 that affected the West Coast of North America was a result of interarea mode instability. In this event, thirty 390MW loads were lost as the Western Interconnection separated into four islands [10] [11].
Prior to the disturbance, hot weather contributed to the high system loading and there was a high transfer level from the hydro generation in Canada and the Northwest to California. However, the transfer limits were not violating security constraints.

A number of faults occurred on the 500kV network that resulted in a low voltage problem in sections of the grid. In response, the reactive power output at the McNary

plant in Oregon was boosted from 260MVAr to 490MVAr. The McNary plant comprised 13 units, which tripped one by one due to a protection error. When the plant power output had dropped to about half of its original value, oscillations at 0.224Hz became negatively damped. These continued to grow to 1000MW peak-to-peak, tripping the three lines comprising the California-Oregon intertie, which had been carrying about 4350MW prior to the disturbance and caused the first separation. This resulted in voltage and frequency disturbances that separated the system into four electrical islands and caused widespread blackouts.

It is worth noting that the system model produced a stable response in the McNary unit trips in the initial reconstruction of the event as shown in figure 3.2. After considerable effort to improve the system dynamic model, the negatively damped response was reproduced.

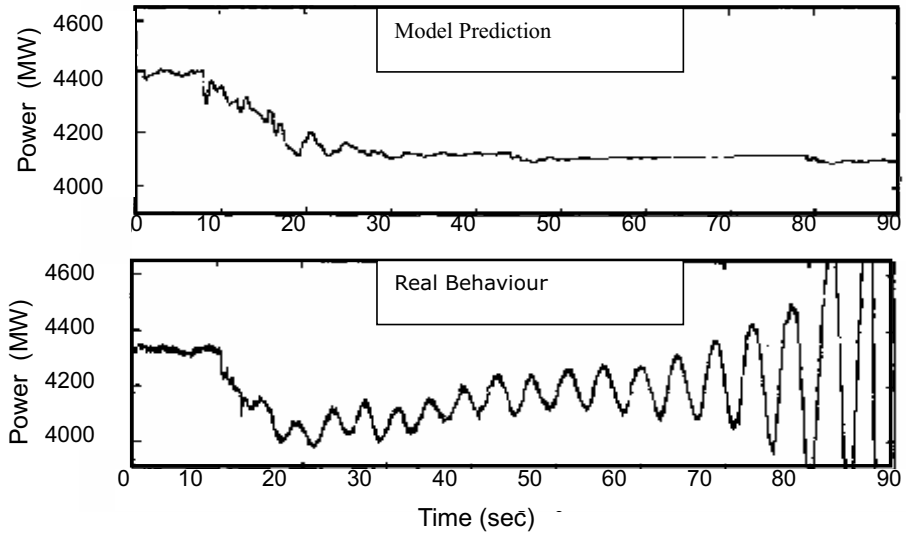


Figure 3.2: Model Prediction and Actual System Behaviour

Analysis of the damping record shown in figure 3.3 reveals several warning signs that might have been used to indicate the system was experiencing inadequate damping prior to the onset of the negatively damped oscillations. In a comprehensive review of the August 1996 event, Venkatasubramanian describes the causes in terms of linear and non-linear dynamics [12]. The author describes a number of features of the event:

- Prior to the blackout, the system topology was weakened and the system was N-1 insecure as any one of a number of events could have resulted in dynamic instability.
- The low voltage profile at the time had degraded the damping of the interarea mode
- High power flow from north to south in the interconnection affected the damping
- The initial period of small-disturbance instability resulted in large but stable oscillations due to limit cycling
- After a period of stable limit cycling, a further shift in behaviour resulted in unstable limit cycling. This led to very large oscillations that caused the system collapse

The last two points illustrate that the behaviour of the system during unstable conditions is unpredictable. An unstable condition may be contained or out of control and it is not practically possible to predict how the system would respond prior to an event.

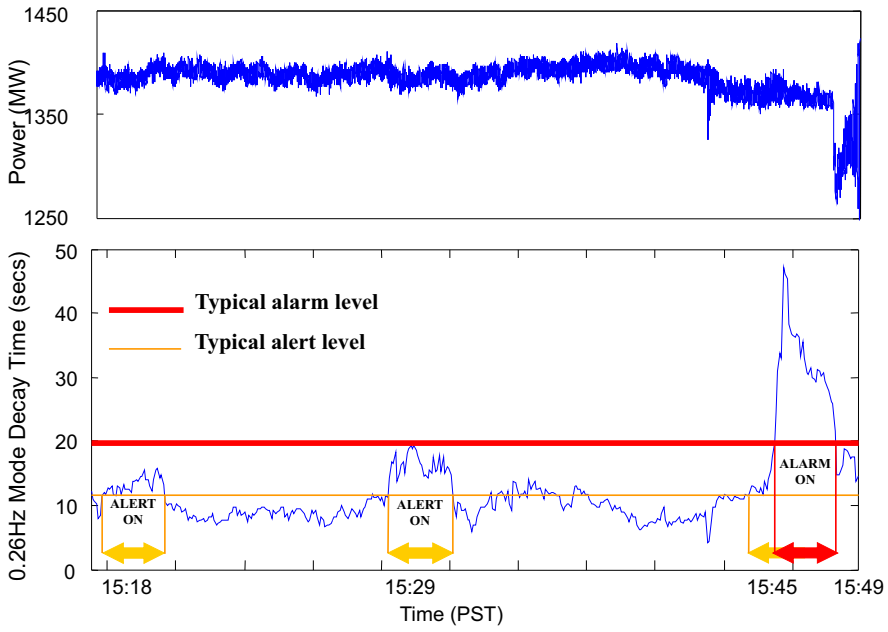


Figure 3.3: 0.26Hz Mode decay time constant leading to West USA/Canada System Separation, 1996

3.2.3 Scandanavia, January 1997

An example of an unstable oscillation resulted in an outage of two nuclear units in Sweden. Prior to the disturbance 1 January 1997, the Nordel system load was high due to the winter heating load. Power was being imported from Denmark and wheeled through Sweden into Norway [13].

A busbar earth fault occurred and was cleared in 60ms. The fault resulted in disconnection of the lines connected to the faulted busbar. At this stage, power oscillations occurred but these were positively damped. However, four seconds after the initial fault, a second line tripped on overload and the oscillations became negatively damped. The line outages had resulted in a weak interconnection to Ringhals nuclear units 1&2 through two relatively long radial lines. When these lines were tripped, the power oscillations became positively damped and the system was stabilized. The frequency of the oscillation was approximately 0.5Hz. The nuclear units were subjected to an emergency shutdown and 1700MW of generation was lost. The system frequency dropped to 49.4Hz but emergency procedures worked correctly and the incident was contained without system-wide blackout.

Large power oscillations were also observed in the widespread Scandanavian blackout of 2003 [14]. In this case, there was a double busbar fault at a substation connecting Ringhals units 3&4, shortly after a unit trip at another nuclear plant. Due to the busbar fault, Ringhals units 3&4 tripped and large power oscillations were observed at units 1&2. In this case, however, the power oscillations were positively damped and reduced over a period of 90 seconds. This indicates that the power oscillations were close to instability and positively (though poorly) damped. The subsequent events that caused the eventual blackout were not related to oscillatory instability.

3.2.4 Further Comments on the Case Studies

A number of general points can be drawn from the above case studies:

- Classical instability can occur with no triggering event, e.g UK Case Study

- Stability may be sensitive to power flow, e.g UK, West USA, Canada
- Excitation systems contribute positively or negatively. In the UK case, the problem was resolved by installation of power system stabilizers
- Stability is dependent on the network topology and impedance, e.g West USA/Canada, Scandanavia
- Stability is sensitive to the network voltage profile, e.g West USA/Canada

It is worth noting that in a number of case studies, the unstable condition was not replicated by post-event analysis using the existing dynamic model. The condition was replicated by modeling, only after a validation exercise was carried out with the knowledge of the actual system response.

3.3 Source Location Methods from Literature

3.3.1 Feature Selection Methods

Large power systems can be in a large number of system states. These can be described by such measurable values as bus voltage, voltage angle, real and reactive power flows as well as generated supply and demand. The information may also include topological data such as transformer settings, switch positions and system topology. For large interconnected systems, the system state information is too large to use any effective computer intelligence methods such as neural networks or decision trees to solve oscillatory stability problems [15].

Therefore, the data must be reduced before it can be used in any of the computer intelligence (CI) methods. According to the literature, the system input variables may be classified as either attributes or features. In general, a reduced set of features must represent the entire system, since a loss of information in the reduced set results in a loss in performance and accuracy in the CI methods. In large interconnected power systems, it is difficult to extract relationships between features and targeted oscillatory stability. This is because the system is highly non-linear and complex. For this reason, feature selection cannot be performed by engineering judgment or physical knowledge

alone, but must be implemented according to the statistical property of the various features and the dependency among them.

In the literature, there are many different approaches to data reduction in power system security and stability applications. The most common methods are the feature selection and feature extraction techniques [16] [17] [18] [19]. The typical selection and extraction methods which can be studied in detail in [15] and [20], are:

- Similarity (Correlation, Canonical Correlation)
- Discriminant Analysis (Fisher Distance, Nearest Neighbour, Maximum Likelihood)
- Sequential Ranking Methods
- Entropy (Mutual Information, Probability)
- Decision Tree Growing
- Clustering
- Principal Component Analysis

Although there are many different approaches, all of them are based on either distance or density computation.

3.3.2 Selection Approaches

In this section of the literature review, three different selection approaches for feature reduction are compared. The first approach applies principal component analysis (PCA), followed by a selection based on clusters created by the k-means cluster algorithm. The second approach is based on the decision tree (DT) criteria for the separability of large data sets and the third approach implements a genetic algorithm (GA) for selection based on DT criteria. The complete selection procedure is shown in figure 3.4. Before the approaches are applied, the initial feature is pre-selected by engineering judgment and the data is normalized.

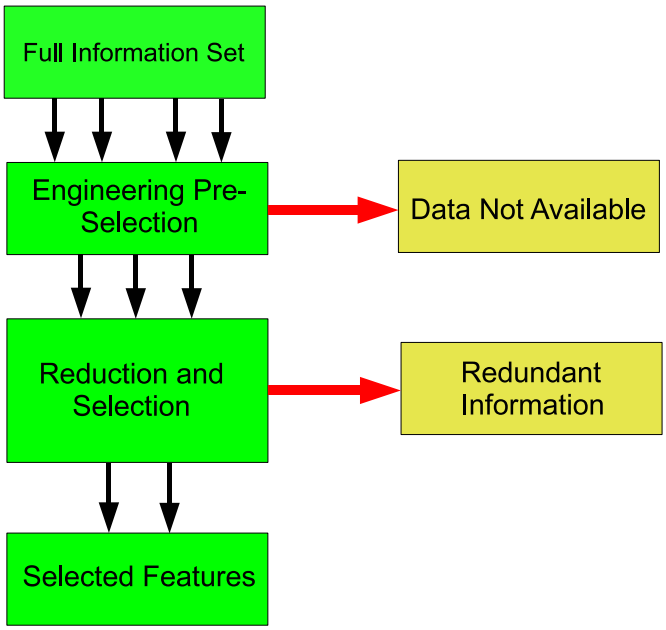


Figure 3.4: Basic Concept for Feature Selection Procedure

3.3.3 Engineering Pre-Selection

The idea of pre-selection on the raw data is based on engineering judgment. It is necessary to collect as much data as possible from the system which can be used in a stability assessment. The focus here is on features that are both measurable in the actual power system and available from the power utilities.

The features used in this literature review are generator related such as real and reactive power generation for each machine as well as their summation per area. In addition, the stored kinetic energy is used and is defined as installed MVA of running blocks multiplied by the inertia constant. The stored kinetic energy is computed for both individual machines and all machines in the same area. Moreover, the real and reactive power on all transmission lines, voltages and voltage angles on all node buses are used as features because they may contain important information about the loadflow in the system, especially if the system is large.

3.3.4 Normalization

In [21], the data is subjected to normalization before it is used in any subsequent numerical steps. Normalization is a transformation of each feature in the dataset.

Data can be transformed either to the range [0,1] or [-1,1], or can be normalized to obtain a zero mean with unit variance which is the most applicable way according to the literature.

The standardized score is computed by the deviation feature vector from its mean, normalized by its standard deviation. The z-score called $\mathbf{z}_j$, is computed for every feature vector $\mathbf{x}_j$, including p patterns given by:

$$z_j = \frac{x_j - \bar{x_j}}{\sigma_x} \tag{3.1}$$

with the standard deviation of feature vector $\mathbf{x}_j$:

$$\sigma_x = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(x_i - \bar{x})^2} \tag{3.2}$$

and the mean of feature vector $\mathbf{x}_j$:

$$\bar{x} = \frac{1}{p}\sum_{i=1}^{p} x_i \tag{3.3}$$

3.4 Selection Approach I

In the first step, the data is reduced in dimension by the principal component analysis (PCA), which is characterized by a high reduction rate with a minimal loss of information. The PCA is fast and can be applied to large datasets [22]. The PCA projections onto lower dimensional space determine which features are projected onto the principal axes. This information can be used for selection since similar features are projected onto the same principal axes. Therefore, the projections rather than the features are clustered on lower dimensional space using k-means clustering algorithm. The dimensionality reduction before clustering is necessary when processing large amounts of data because the k-means cluster algorithm leads to the best results for small and medium sized data sets.

3.4.1 Reduction by PCA

Let **F** be a normalized feature matrix of dimension $p \times n$, where n is the number of original feature vectors and p is the number of patterns. The empirical covariance matrix **C** of the normalized **F** is computed by:

$$C = \frac{1}{p-1} \cdot F^T \cdot F \tag{3.4}$$

Let **T** be a $n \times n$ matrix of eigenvectors of **C** and the diagonal variance matrix Σ^2 is given by:

$$\Sigma^2 = T^T \cdot C \cdot T \tag{3.5}$$

Σ^2 includes the variances σ_x^2. Notice that the eigenvalues λ_k of the empirical covariance matrix **C** are equal to the elements of the variance matrix Σ^2. The standard deviation σ_k is also the singular value of **F**:

$$\sigma_k^2 = \lambda_k \qquad (1 \leq k \leq n) \tag{3.6}$$

The n eigenvalues of **C** can be determined and sorted in ascending order $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$. While **T** is a n-dimensional matrix whose columns are the eigenvectors of **C**, $\mathbf{T}_q$ is a $n \times q$ matrix including q eigenvectors of **C** corresponding to the q largest eigenvalues of **C**. The value of q determines the size of the new dimension of the features and is smaller than n. It also determines the retained variability of the values, which is the ratio between the first q eigenvalues and the sum of all n eigenvalues [22].

$$\upsilon = \frac{\sum_{i=1}^{q} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{3.7}$$

From the basic equation of PC transformation:

$$F^{(PC)} = F \cdot T \tag{3.8}$$

follows:

$$F \approx F_q^{(PC)} \cdot T_q^T \tag{3.9}$$

where F^{PC} contains the selected PC feature vectors. According to equation (3.9), T_q can be interpreted as a loading matrix and F^{PC} as a not normalized factor matrix. The idea is to use columns of T_q^T instead of the high dimensional original feature vectors $\mathbf{F}$ for clustering.

3.4.2 Clustering by k-means

The n columns of the T_q^T vectors, which represent the projection of the ith feature $\mathbf{F}$ onto a lower q-dimensional space. Therefore, the k-means cluster algorithm [23] shows with T_q^T a better performance and accuracy when applied to the original features directly. This is why both techniques, PCA and clustering, are used in combination. Nevertheless, it is even possible to skip the PCA computation altogether, which means that the original features are clustered from the beginning.

Once the columns of T_q^T have been computed, the k-means algorithm clusters them into k-groups, whereby k is independent of q. Because of the similarities between the features within a cluster, one can be selected and the others can be treated as redundant information. The feature in one cluster, which is closest to the centre of this cluster can be chosen. This group of k features will be maintained. Hereby, the centre $\mathbf{c}$ of a cluster including m vectors $\mathbf{a}_i$, is defined as:

$$c = \frac{1}{m} \cdot \sum_{i=1}^{m} a_i \tag{3.10}$$

and the distance d between the center and the vector $\mathbf{a}_i$ is computed by:

$$d(a, c) = \|c - a\| \tag{3.11}$$

The remaining variability provides information about the size to which the full set can be reduced without any noticeable loss of information.

In the literature this technique was applied to the 16 machine case which is used later in this thesis in chapter 5. The total number of features associated with the 16 machine model is 301. In the reduced study, 50 features represent the full feature set and

remarkably 10 features represent 95% of the variability.

For applicable online stability assessment, the number of features to be selected also depends strongly on the power system size. On the one hand, the redundancy must be reduced to a minimum for successful CI application. On the other hand, the redundancy must not be eliminated completely in case of missing data. If there is still some redundancy in the information, it is possible to restore missing data from the remaining correct input data.

3.4.3 Final Feature Selection

The process of data clustering must be followed by a final selection to obtain the input features for the OSA methods. The k-means cluster algorithm provides a feature ranking based on the distance between the cluster centre and the features within the cluster. When used automatically, the algorithm selects the feature from one cluster, which is close to the centre of the cluster. This way of selection is based on a mathematical relationship and does not include any physical or engineering intention. To include engineering knowledge in the selection process, the features inside a cluster are treated as physical measurements from the power system and judged by criteria such as technical measurability, availability from the utilities and expected usefulness of CI based OSA.

For example, the transmitted real power between two network areas might be a much more useful input feature than a voltage angle on a bus. The voltage at a bus might be more applicable than the generated real power on a particular generator because the voltage is both measured and shared.

3.5 Selection Approach II

The last section introduced a cluster algorithm for the selection of CI input features. However, the process of clustering is lacking in the fact that it is not concerned with the CI target. The clustering algorithm leads to a result depending on the natural structure of the presented input data. However, this result is not impacted by the

power system stability assessment problem. In other words, a certain feature showing no correlation with other features may be selected in a single cluster, but this feature is not necessarily important for OSA.

In order to improve the result, another way of feature selection is introduced in this section. It is based on decision trees (DT's) and the key idea is to grow a DT with a full input data set. The tree is grown with an algorithm, which determines the tree nodes of highest data separability and places them on top of the tree. In other words, the top nodes of the tree give the most important information about how to separate the data set and can therefore be used as input features. The DT is grown with the entire data set as inputs and features corresponding to the k nodes from the top, are selected as input features [15].

3.5.1 Decision Trees

DT techniques belong to CI methods and became highly popular in the age of modern computers. They are based on a sequence of questions that can either be answered yes or no. Each question asks whether a predictor satisfies a given condition, whereby the condition can both be continuous and discrete. Depending on the answer to each question, one can either proceed to another question or arrive at a response value. DTs can be used for non-linear regression (regression trees) when using continuous variables, or they can be used for classification (classification tree) when using discrete classes. When used for feature selection, the DT is grown as a regression tree [24].

Without prior knowledge of the non-linearity, the regression tree is capable of approximating any non-linear relationship using a set of linear models. Although regression trees are interpretable representations of a non-linear input/output relationship, the discontinuity at the decision boundaries is unnatural and brings undesired effects to the overall regression and generalization of the problem [25].

3.5.2 Tree Growing

To construct a tree, the data is split into two sets. One set is used to learn the tree and the other set is used to test it afterwards. For tree growing, there are different

algorithms available depending on the kind of tree desired. Regardless of the algorithm, the first task is to find the root node for the tree. The route node is the first node splitting the entire dataset into two parts. Therefore, the route must do the best job in separating the data. The initial split at the route creates two new nodes called branch nodes. The algorithm searches at both branch nodes again for the best split to separate the subsets.

Following this recursive procedure, the algorithm proceeds to split all branch nodes by exhaustive search until either a branch node contains only patterns of one kind, or the diversity cannot be increased by splitting the node. The nodes, where the tree is not further split, are labeled as leaf nodes. When the entire tree is split until only leaf nodes remain, the final tree is obtained [24] [26] [27].

3.5.3 Growing a classification tree

In classification, the algorithm splits the patterns in such a way, that each branch node t performs best in splitting the data into separate classes. An error measure $E(t)$ is defined to quantify the performance at a node t. It is also referred to as the impurity function ϕ describing the impurity of the data or classes under a given node. If all patterns belong to the same class, the impurity function attains a minimum value of zero. If the patterns are equally distributed over all possible classes, the impurity function will reach a maximum value.

When p_j is the percentage or probability of cases in node t that belong to class j in a J-class problem, the error or impurity measure $E(t)$ of a node t is computed by the impurity function ϕ in node t by:

$$E(t) = \phi(p_1, p_2, \ldots, p_J) \tag{3.12}$$

with:

$$\sum_{J}^{j=1} p_j = 1 \quad and \quad 0 \leq p_j \leq 1 \tag{3.13}$$

and the impurity measure for the complete tree is computed by the summation of all branch node impurities.

However, the impurity function ϕ can be computed in different ways, the best known functions for classification are the *Entropy Function* equation and the *Gini Diversity Index* equation shown in equations (3.12) and (3.13) respectively. The *Entropy Function* is also known as the *Maximum Deviance Reduction*. Both functions are always positive unless all patterns belong to the same class. In this case, the functions obtain zero impurity. They obtain their maximum values when the patterns are equally distributed over all possible cases.

$$\phi = -\sum_{J}^{j=1} p_j \cdot ln(p_j) \tag{3.14}$$

$$\phi = 1 - \sum_{J}^{j=1} p_j^2 \tag{3.15}$$

3.5.4 Growing a regression tree

When growing a regression tree, a local model is employed to fit the tree at node t to the dataset at node t. Analogous to most regression methods, the error measure is implemented by the least squares approach [28]. The mean squared error or residual at node t is computed by:

$$E(t) = \frac{1}{N_t} \cdot \underbrace{min}_{\theta} \sum_{i-1}^{N_t} (y_i - d_t(x_i, \theta))^2 \tag{3.16}$$

where $\mathbf{x}_i$ is the input vector. The corresponding output target is y_i and N_i is the number of patterns at node t. The local model at node t is given by $d_t(\mathbf{x},\theta)$, whereby θ is a vector which can be modified. If $d_t(\mathbf{x},\theta) = \theta$ is a constant function independent of the input vector $\mathbf{x}_i$, the local model $d_t(\mathbf{x},\theta)$ at node t is computed by the mean value of the target y_i which is:

$$\bar{y_t} = \frac{1}{N_t} \cdot \sum_{i=1}^{N_t} y_i \tag{3.17}$$

since any subset of y will minimize equation (3.16) and (3.17) for the mean of y as a local model and thus:

$$E(t) = \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i - \bar{Y_t})^2 \tag{3.18}$$

If $d_t(\mathbf{x},\theta)$ is a linear model with linear parameters θ, a least squares method is applied to identify the minimizing parameters $\bar{y}$ [24] [25].

According to the chosen model, the tree is split in order to maximize the error decrease for the complete tree. The process of learning is fast and applied once in the beginning. When the tree is grown, it can be used to predict an output depending on the presented inputs. It has been observed that large decision trees usually do not retain their accuracy over the whole space of instances and therefore tree pruning is highly recommended [29].

3.5.5 Tree Pruning

Pruning is the process of reducing a tree by turning some branch nodes into leaf nodes and removing the leaf nodes under the original branch. Since less reliable branches are removed, the pruned DT often gives better results over the whole instance space even though it will have higher error over the training set. To prune a tree, the training set can be split into the growing set (for learning the tree) and the pruning set (for tree pruning). Different pruning approaches use the testing procedure for pruning. However, pruning is necessary to improve the tree capability and reduce the error cost. Moreover, large trees become specific in relation to the used growing data and some lower branches might be affected by outliers.

Pruning is basically an estimating problem. The best tree size is estimated based on the error cost. When only training data is used to prune, the estimation is called *internal estimation*. Accuracy is computed by counting the misclassification at all tree nodes. Then the tree is pruned by computing the estimates following the bottom-up approach (post-pruning). The resubstituting estimate of the error variance for this tree and a sequence of simpler trees are computed. Because this estimation probably

underestimates the true error variance, the cross-validated estimation is computed next. The cross-validation estimate provides an estimate of the pruning level needed to achieve the best tree size. Finally, the best tree is the one that has a residual variance that is no more than one standard error above the minimum values along the cross-validation line [24].

3.5.6 Final Feature Selection

For the selection purposes the DT is grown with the entire dataset as an input. There is no need for a testing set because the focus is on the branch splitting criteria and not the tree accuracy. The DT is pruned to a level where only the desired number of features remain. Hereby, the repeated selection of the same feature is prevented. However, the features corresponding to the k-nodes from the top, which do not contain any feature more than once, are selected as input features.

3.6 Selection Approach III

The previous section showed that the DT method allows a fast and accurate way to determine the quality of the feature selection approaches I and II. However, when the selection quality can be evaluated, it is possible to implement a genetic algorithm (GA) to search for the best feature selection. The objective function evaluates the fitness of the GA method. The GA is a global search technique based on the evolutionary process from nature. It is highly applicable in the cases of discrete selection or mixed integer programming. The GA is an optimization algorithm which searches an input space while minimizing an objective function under given constraints [30] [31]. The key idea of a DT based GA is shown in figure 3.5.

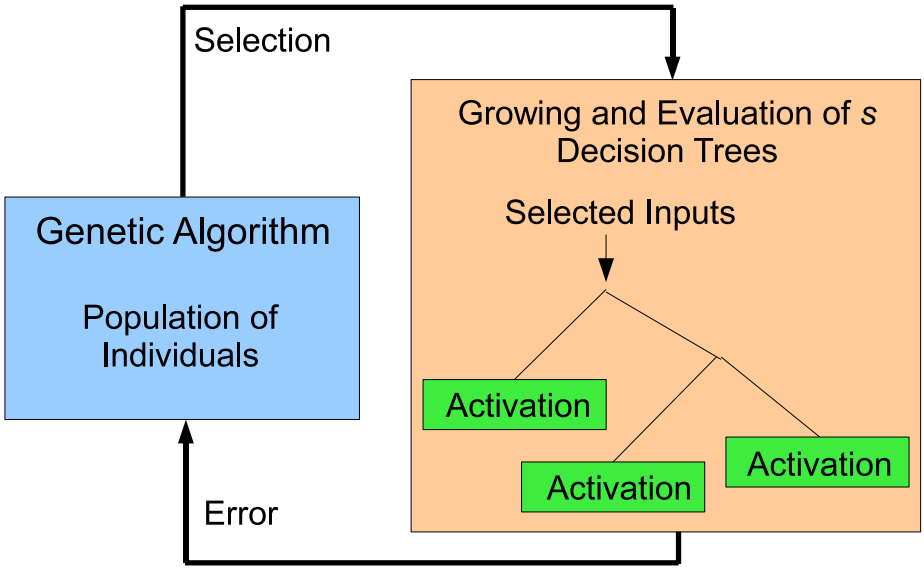


Figure 3.5: Block Diagram for the Feature Selection Process based on GA's

3.7 Genetic Algorithm

The GA involves a population of individuals, which includes coded information. Individuals who perform poorly on a given problem are discarded while more successful ones produce variants of themselves. Based on the Darwinian analogy of "survival of the fittest", the population will improve over time and reach one optimal solution. However, the solution may be found to be a local optimum and is not necessarily a global optimum. In this study, the individuals are of binary type containing zeros and ones. Each individual has one bit for each feature from the total feature set. This is size 252 bits. Any bit showing a *one* means that the corresponding bit is selected.

The entire population consists of many individuals starting at different initial selections. The fitness of those individuals is computed by an objective function based on DT's. Hereby, the objective function grows one DT for each of the sampling points. The DT's are grown with training data and the error evaluated with testing data. When z is the testing target and y the DT output, the sum squared error (SSE) is computed over all p patterns of the testing data set. Then the SSE is accumulated over

all s DT's. Thus the error is computed according to equation (3.19):

$$E = \sum_{j=1}^{s} \sum_{i=1}^{p} (z_{ij} - i_{ij})^2 \tag{3.19}$$

The GA evaluates at each iteration step the fitness or error of all individuals. The individuals showing the best selection and therefore the smallest error are chosen as parents for the next generation.

The next generation is created based on the fitness of individuals and a reproduction and crossover to find the best parents, which will generate the best offspring. The reproduction involves a random process with a particular probability for being chosen as a parent. Crossover is performed on the result and mutation applied randomly to allow discontinuities in the search algorithm, which may lead to significant improvement.

When compared to approach I and II, the advantage of this is that it does not need a pre-defined number of features. Both clustering and DT selection require a predefined number of features to be selected. However, the GA will not only find the best combination but also the number of features leading to the best results.

3.8 Computational Intelligence Methods for Source Location

This section of the literature review presents some of the CI methods used for oscillatory stability assessment (OSA). The inputs are selected as normalized features as discussed earlier in this section. The outputs are assigned to the oscillatory stability problem and its assessment. The chosen CI methods presented here include neural networks (NN's) such as multi-layer feedforward NN's and radial base NN's, DT's in the form of regression and classification trees, and ANFIS.

NN's are chosen for OSA since they are powerful and adaptive tools in many applications [32] [33] [34]. The ANFIS, which is a combination of fuzzy logic and NN, is also

highly applicable and benefits from both techniques [25] [35]. The DT methods are of interest since they are based on simple rules which can be easily understood by the user [24] [27].

3.8.1 Neural Networks

Essentially, a NN consists of a series of connected neurons. The neurons can be considered as nodes, which get input signals from other neurons and pass the information to selected neurons [36] [37]. The connections between the neurons are weighted by factors. A factor of 1 passes the entire information, a factor of 0 passes no information, and a factor of -1 passes the negative of the signal. Usually a neuron adds the incoming input information x_i (weighted output from other neurons) and passes the sum net through an activation function f. The neuron output is the output of the activation function $f(net)$ of this neuron.

Figure 3.6 illustrates the single neuron structure for three inputs coming from three different neurons. The corresponding neuron output is called out and computed by equation (3.20).

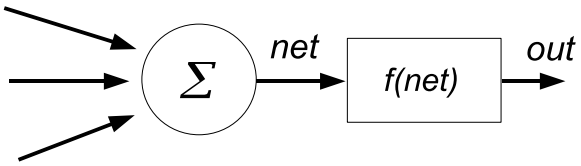


Figure 3.6: Single Neuron Structure

$$out = f\left(\sum_i x_i\right) \tag{3.20}$$

A single neuron as shown in figure 3.6 is not powerful enough to learn complex relationships. But the connection of a group of neurons to a so-called layer can increase the capability of a network tremendously. Further, each neuron can get a bias. This is an additional input, which does not come from another neuron. It is an input that allows a non-zero value for net even when all the inputs x_i are zero. The bias is needed in the network to generalize the solution. The neuron output including a bias b is analogous

to equation (3.20) given by:

$$out = f\left(\sum_i x_i + b\right) \tag{3.21}$$

Activation functions process the summated neuron input before passing it to the output. The basic advantage is to model nonlinear processes by using non-linear activation functions. A sigmoidal hidden layer feeding a linear output layer can represent most processes [39].

Any function $f(x)$ can be used as an activation function, but commonly used functions are those which can be derived at any value of x. The reason activation functions are used is that they implement the backpropagation algorithm for NN training, where the first order derivation of the activation function is needed. A sigmoidal activation function, also called logsig, is given in equation (3.22). Another often implemented activation function is the hyperbolic tangent function in equation (3.23).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.22}$$

$$f(x) = tanh(x) \tag{3.23}$$

Most activation functions give outputs within the intervals [0,1] or [-1,1]. These functions are usually used in combination with other patterns, which are also normalized within the range. The advantage or normalization is that the same range is used for each of the inputs, whereby the entire input range for every input neuron is used. Thus, the NN can develop its maximal capability with normalized data patterns.

The best NN architecture depends on the problem to be solved. For a simple problem, one layer might be enough. If the problem becomes more difficult e.g. in the case of a non-linear relationship between inputs and outputs, the network needs to be expanded. The more neurons in a hidden layer, the more weights can store the information [38]. For a classification problem, a layered perceptron or a radial base NN might be reasonable. The multilayer feed-forward network is the common network type used for approximation of non-linear relationships between input and output.

3.8.2 Multi-Layer Feedforward NN's

The multilayer feedforward network consists of at least three layers: an input layer, a hidden layer and an output layer. The term feedforward means that there is no feedback from one neuron layer to the previous neuron layer. The information is passed only in one way and that is forward. This network structure is commonly used for all kinds of approximation. With the backpropagation algorithm implemented as a training algorithm, it can be very powerful and lead to accurate results regarding pattern approximation of non-linear relationships between input and output. A disadvantage of the feedforward network is that it cannot perform temporal computation.

Once the network architecture is fixed, it needs to be trained. Training is the process of presenting data patterns to the network. The goal of training is that the network stores the relationship between the input and output in its weights. During training, a pattern is presented to the NN. The NN passes the input information to the outputs according to the weights and activation functions in the network. The NN output, which corresponds to the given input pattern, is compared to the corresponding target for this input. The difference between the NN output and the target leads to an error, which is used to update the weights. The weights are updated in a way to minimize the error of each pattern. This is repeated again and again until the error is small enough for the given application.

After training the network is tested. Testing consists of presenting new patterns to the network that have not been used in training. The results of the testing show whether the NN is able to give reasonable answers to new patterns. If the testing is successful, the NN is able to compute outputs to given inputs similar to the trained patterns.

3.8.3 Probabilitistic Neural Networks

Probabilitistic Neural Networks (PNN's) belong to a group of radial basis networks and are often used for classification. The network consists of two layers, the first one is called a radial basis layer and the second one is a competitive layer. The task of the

first layer is to compute the distances between the presented input vectors and their weight vectors. For each of the vectors, the element showing the highest probability is chosen and assigned to the corresponding class.

The PNN design is influenced by the sensitivity of the radial basis function of type $y = exp(-x^2)$. As the presented vector x decreases, the output y of the radial basis function increases. When the distance d between a given input vector and weight vector is weighted with a sensitivity parameter *spread*, the radial basis function is written in the form:

$$f(d) = exp\left(-ln2 \cdot \frac{d^2}{spread^2}\right) \tag{3.24}$$

Since the radial base functions act as detectors for different input vectors, the weight vectors are computed accordingly and there is no need to train the network. The PNN design is therefore straightforward and it does not depend on a training process [39].

3.8.4 Classification of System States

In this particular study, the decision boundary for sufficient/insufficient damping was set at 4%. When a pattern includes no eigenvalues with corresponding damping coefficients below 4%, the load flow of this pattern is considered to have sufficient damping. When at least one of the modes is damped below 4%, the system is considered insufficiently damped.

There are many tools used for classification and this section compares the PNN and the multi-layer feed-forward NN. Both NN's are implemented with the same set of normalized patterns and the same patterns for training and testing.

3.8.5 Classification by PNN

In the PNN case, when the spread is large the training errors tend to be large. When the spread is small, the training errors tend to be small with testing errors being large. A medium spread of 0.05 produces both small training and testing errors.

However, when a PNN is used the patterns are classified as small errors below 1.5%. Moreover, most of the error patterns belong to eigenvalues with a damping coefficient

close to 4%. Since this is a classification border, most occur near 4%. When neglecting these patterns, which are still sufficiently damped, the classification error for the remaining patterns will be much less than 1.5%.

3.8.6 Classification by Multi-Layer Feed-Forward PNN

The classification can also be performed by a common feed-forward multilayer NN with one hidden layer. The inputs are selected features from the power system and the output is a single variable. For sufficient damping, the target is 0 and for insufficient damping the target is 1. The classification decision is made at 0.5 with errors resulting mostly from these patterns, which include eigenvalues close to the damping boundary of 4%.

When the output is 0 and is therefore below a certain level, the probability of a sufficiently damped situation is high. When the output is close to 1, the probability for an insufficiently damped situation is very high. When the output is somewhere between the two limits, the classification is unidentifiable. These classifications have a weakly damped mode around 4% and can be classified by other OSA methods.

3.8.7 Eigenvalue Region Classification

Instead of classification between sufficiently and insufficiently damped patterns, the observation area can be divided into small regions and the existence of dominant eigenvalues inside these regions can be identified.
The observation area in the complex plane is split into 12 regions corresponding to a damping range from -2.5% to 5.5% and a frequency range from 0.22Hz to 0.84Hz. In each of these regions a separate PNN for classification is used. The regions overlap slightly to reduce the false dismissals at the classification borders. The overall errors depend on the number of used regions and the tolerance for overlapping. When the overlapped region is increased, the false dismissals are reduced because the overlapping guarantees that at least one of the regions detect the eigenvalue. However, an optimum overlapping size exists but false alarms cannot be eliminated completely due to the need for an acceptable resolution of eigenvalues and hence a reasonable number of

areas.

When the number of regions are increased, the accuracy of the regions are also increased. However, the errors are more numerous so using a large number of regions is not recommended.

3.8.8 Eigenvalue Region Prediction

In this study taken from [21], the damping region is chosen to be 4% and -2%. Thus the observation area is sampled along the real axis (σ) in the range of damping coefficient between 4% and -2%. This is done five times for five different frequencies f. The width of one σ sampling step is $\Delta\sigma$, the width of one f sampling step is Δf. These step widths are constant for the entire sampling process:

$$\Delta\sigma = 0.0216s^{-1} \qquad \Delta f = 0.15Hz \tag{3.25}$$

The sampling in this case results in 47 sampling points, each row of which has its own NN. It is usual for the samples to be spread over the entire eigenvalue space. This is done because the NN's are not high precision tools and cannot be accurate in narrow zones. Once the NN's are trained properly, the NN results are transformed from activation level into eigenvalue locations and regions respectively. This transformation is done using sampling point activation. For all patterns, the activation values given from the sampling points are used to setup an activation surface, which is used to construct a region in which the eigenvalue is located.

For error evaluation, the predicted regions are compared with the positions of the eigenvalues. If the eigenvalue is located inside a region then the prediction is correct. If an eigenvalue is located outside the corresponding region or no region is constructed for this eigenvalue, the error is false dismissal. If a region is constructed, but no corresponding value exists within the observation area, the error is false alarm. The errors shown in table 3.1 are the sum of the false dismissal errors and the false alarms.

Inputs	Selection Algorithm	Learning	Testing
5	KM-05	11.28%	10.95%
10	KM-10	5.96%	7.10%
16	GA-16	6.95%	7.40%
20	KM-20	3.87%	3.55%
30	DT-30	3.41%	2.66%
40	DT-40	3.01%	3.25%
50	DT-50	3.34%	2.96%

Table 3.1: Eigenvalue Region Learning and Testing Misclassification Percentages by NN training for different input dimensions

When the number of inputs is 20 or above, the errors are below 4% but for fewer inputs, the errors increase considerably. Note that errors depend highly on NN architecture, the parameter settings and the training process. As a comparison, the accuracy of the predictions are compared. The mean and the standard deviation of the damping difference $\Delta\zeta_{reg}$ are computed for all correct predicted patterns, where the eigenvalues are inside their predicted areas. The mean and standard deviation are shown in table 3.2. The mean and standard deviation of the frequency difference Δf_{reg} are computed similarly and listed in table 3.3.

	Training		Testing	
Inputs	Mean	**STD**	Mean	**STD**
5	1.181	0.823%	1.149%	0.786%
10	1.176	0.804%	1.131%	0.758%
16	1.188	0.790%	1.164%	0.764%
20	1.182	0.783%	1.164%	0.759%
30	1.197	0.796%	1.177%	0.758%
40	1.185	0.786%	1.148%	0.768%
50	1.196	0.806%	1.170%	0.759%

Table 3.2: Mean and STD for $\Delta\zeta_{reg}$ on %

	Training		Testing	
Inputs	Mean	**STD**	Mean	**STD**
5	0.098	0.055%	0.096%	0.053%
10	0.105	0.054%	0.100%	0.053%
16	0.105	0.054%	0.102%	0.051%
20	0.108	0.055%	0.106%	0.051%
30	0.111	0.052%	0.109%	0.049%
40	0.109	0.053%	0.103%	0.051%
50	0.110	0.054%	0.108%	0.052%

Table 3.3: Mean and STD for $\Delta\zeta_{reg}$ on %

3.9 Neuro-Fuzzy Networks

Neuro-Fuzzy methods are a combination of NN and Fuzzy Logic systems. NF methods can be very powerful in certain applications. Fuzzy logic systems need a set of rules to be pre-defined and one might not have knowledge of these rules. However, automated parameter tuning by a NN structure inside a fuzzy logic system may be used to replace the a priori knowledge of the power system.

A typical fuzzy inference system (FIS) maps a given input to an output on the basis of fuzzy rules and membership functions, which are chosen arbitrarily. The disadvantage of the FIS is the possible lack of knowledge about some rules. Every FIS consists of the following components:

- Fuzzyfication by defined membership functions
- Fuzzy Inference Mechanism
- Defuzzyfication

Fuzzyfication is the process of converting crisp statements into fuzzy statements. This is done with the use of membership functions (MF) which determines the extent to which a variable belongs to a pre-defined class. This indicates a concept called *degree of membership*. Commonly, MF's are described by simple geometric functions such

as triangular or trapezoidal functions. More smooth MF's use Gaussian functions [40] [41].

Next, the fuzzyfied premises are evaluated by rules which include the knowledge of FIS. The rules are predefined and depend both on the nature of the application and the experience of the designer. The challenge is to find the rules, which best describe the behaviour of a given system. Typical fuzzy inference mechanisms have been developed by Mamdami and Tsukamoto [42] in the form of:

$$if \quad x_1 = A_{1i} \quad AND \quad x_2 = A_{2i} \quad THEN \quad y = B_i \tag{3.26}$$

Since the output does not depend directly on the input, Takagi and Sugeno proposed their fuzzy mechanism as:

$$if \quad x_1 = A_{1i} \quad AND \quad x_2 = A_{2i} \quad THEN \quad y = f(x_1, x_2) \tag{3.27}$$

whereby the output is a function dependent on the inputs. In the case of a constant function f, the system is called a zeroth order system. When f is a linear function, the system is of the first order [43]. When applying those using constants or linear functions, it is possible to use optimization techniques to find the best parameters to fit data instead of attempting to do it heuristically.
Finally, the applied rules lead to consequents which are transformed back into crisp consequents. This is called defuzzyfication and in the literature there are many approaches to this end.

3.9.1 ANFIS

When there is no predetermined model structure describing the system, the Takagi-Sugeno FIS can be expanded to a Neuro-Fuzzy system. In the literature the FIS is adapted into an ANFIS (Adaptive NN Based Fuzzy Inference System) as it is more effective for system identification. A Takagi-Sugeno FIS can be generated by one of the following methods:

- ANFIS (Adaptive NN Based Fuzzy Inference System)
- genfis2 (Cluster Based generation of FIS)
- RENO (Regularized Numerical Optimization)

whereby the ANFIS structure is most applicable [44]. The ANFIS defines a Takagi-Sugeno FIS through a NN approach by defining five layers.

ANFIS can be seen as a generalization of CART since it creates a fuzzy decision tree to classify the data into linear regression models to minimize the sum squared errors. But different from decision tree methods, ANFIS is a continuous model [25]. A typical ANFIS structure for two inputs, each with two membership functions is used in this study. The five layers of the ANFIS are connected by weights. The first layer is called the input layer, it receives input data which is mapped onto membership functions. These membership functions determine the membership of a given input to the corresponding class.
The second layer of neurons represents associations between input and output, which are called fuzzy rules. In the third layer, the outputs are normalized and passed onto the fourth layer. Based on the predetermined rules, the output data is mapped in the fourth layer to output membership functions. These outputs are added in the fifth layer and result in a final single valued output.
The ANFIS has the constraint that it only supports Sugeno-type systems of first or zeroth order. The system can only be designed as single output and must be of unity weight for each rule.

3.9.2 Estimation of Power System Minimum Damping

For damping estimation, the ANFIS output is associated with the minimum damping coefficient for the presented load flow scenario. Once the system has been modelled, a part of the input/output datasets (testing set) on which ANFIS was not trained, are presented to the trained ANFIS model, to see how well the trained ANFIS predicts the corresponding output values. It is possible to use another dataset for model validation in ANFIS. This is called a checking dataset and is used to control the potential of the

model fitting data.
When checking data is presented to ANFIS as well as training data, the ANFIS model is selected to have parameters associated with the minimum checking data model error. In this study, 70% of the data was used for training, 20% for checking and 10% for testing.

The reason for using a checking dataset for model validation is that after a certain point in the training process, the model begins to overfit the training data. In principle, the model error for the checking data set tends to decrease as the training takes place, up to a point that overfitting begins. This is when the model error for the checking data suddenly increases.

In this study, the ANFIS was trained with a small number of features i.e. 5-10 inputs from the power system. When trained with a high number of inputs, the system became drastically large in dimension and the training process required both a large amount of time and memory.

3.9.3 ANFIS Results

For evaluation of the trained ANFIS structure, the mean error and the standard deviation of the differences between ANFIS outputs and targets are computed. When done for various ANFIS structures, the best parameter settings can be determined.
The prediction error is defined as the difference between the damping percentage of the ANFIS output and the real damping value. The mean of the errors is nearly zero and therefore only the standard deviations are listed in table 3.4 for training and testing under varying input numbers.

	ANFIS 1		ANFIS 2		
Inputs	Selection	Training	Testing	Training	Testing
5	KM-05	0.28%	0.30%	0.20%	0.32%
6	KM-06	0.23%	0.28%	0.17%	0.17%
7	KM-07	0.21%	0.24%	0.16%	0.17%
8	KM-08	0.18%	0.22%	0.14%	0.16%
9	KM-09	-	-	0.18%	0.18%
10	KM-10	-	-	0.17%	0.17%

Table 3.4: Mean and STD for $\Delta\zeta_{reg}$ on %

When an ANFIS is trained with a large number of inputs, which is above four or five, the time necessary to obtain results is extremely long. The number of inputs and the number of membership functions determine the number of fuzzy rules and therefore training time. Subtractive clustering is a very effective method for reducing the number of rules. Hereby, not only the training time is reduced but the results obtained are also better. The output is accurate enough for online OSA but it gives only sparse information about the power system state. The ANFIS can only be used with one output value that is associated with the minimum system damping. The ANFIS cannot provide any information about the frequency or the number of critical inter-area modes.

3.10 Decision Trees

The DT methods were already introduced in section 3.5.1 when applied to the feature selection. At this point the DT method is used for OSA in the form of classification and regression trees.

3.10.1 Two-Class Classification of System States

The DT method is applied to classify the data into sufficiently and insufficiently damped cases. After learning the tree, it is tested with a different dataset. About 90% of the data is used for growing and 10% for testing. If the tree output and system state

match each other, the classification is correct. Otherwise, the classification is false. The percentage of misclassifications for learning and testing is shown in table 3.5 for varying numbers of inputs.

Inputs	Selection	Learning	Testing
5	KM-05	0.25%	0.93%
10	KM-10	0.31%	0.93%
16	KM-16	0.23%	1.49%
20	KM-20	0.17%	0.93%
30	KM-30	0.25%	1.12%
40	KM-40	0.21%	0.93%
50	KM-50	0.25%	0.75%

Table 3.5: Classification results by decision tree growing for 2 classes and different input dimensions

It is noteworthy that the classification errors in table 3.5 show no strong dependency on the number of inputs.

3.10.2 Multi-Class Classification of System States

In the next step, the two-class classifier is expanded to a multiple class to increase the accuracy of the prediction. Hereby, the DT outputs are assigned to the damping ratios of the least damped eigenvalues. For all computed patterns, the minimum damping coefficient is in the range between -3% to 8%. When the exact damping value is rounded to an integer, the patterns can be distributed into twelve different classes of 1% damping ratio width.

After defining the classes, the DT is grown to assign each pattern to one of these classes. The error distribution for learning and testing is computed and shown on table 3.6 for varying input numbers.

Inputs	Selection	Err. Dist	Learn Err.	Test Err.
5	KM-05	± 1 Class	2.07%	10.07%
		± 2 Classes	0.29%	0.56%
		± n Classes	0.08%	0.56%
10	KM-10	± 1 Class	2.30%	10.07%
		± 2 Classes	0.29%	0.37%
		± n Classes	0.08%	0.00%
16	GA-16	± 1 Class	2.09%	9.14%
		± 2 Classes	0.15%	0.37%
		± n Classes	0.15%	0.75%
20	KM-20	± 1 Class	1.60%	9.70%
		± 2 Classes	0.17%	0.37%
		± n Classes	0.06%	0.56%
30	DT-30	± 1 Class	1.43%	9.70%
		± 2 Classes	0.23%	0.19%
		± n Classes	0.10%	0.00%
40	DT-40	± 1 Class	1.93%	9.89%
		± 2 Classes	0.19%	1.49%
		± n Classes	0.23%	0.19%
50	DT-50	± 1 Class	1.51%	10.26%
		± 2 Classes	0.12%	0.75%
		± n Classes	0.12%	0.00%

Table 3.6: Classification results by decision tree growing for 12 classes and different input dimensions

The errors are divided into different categories depending on the degree of the classifier failure. If the classifier fails but gives a neighbouring class, the prediction is still accurate enough for OSA. But when a class far away from the real one is predicted, the result is useless.

3.10.3 Estimation of Power System Minimum Damping

The final process is for the decision tree to be implemented as a regression tree. The error is defined as the difference between the predicted damping percentage of the tree output and the computed minimum-damping coefficient. The mean of the errors is nearly zero and therefore only the standard deviations are listed in table 3.7 for learning and testing under varying input numbers.

Inputs	Selection	Learning	Testing
5	KM-05	0.11%	0.20%
10	KM-10	0.12%	0.38%
16	KM-16	0.10%	0.25%
20	KM-20	0.10%	0.23%
30	KM-30	0.09%	0.22%
40	KM-40	0.09%	0.23%
50	KM-50	0.09%	0.23%

Table 3.7: STD of differences between predicted and computed damping coefficients for regression tree growing under different input dimensions

The results show small standard deviations of the difference between decision tree outputs and real damping values.

3.10.4 Eigenvalue Region Prediction

The eigenvalue region prediction method as introduced earlier requires multiple CI outputs to predict multiple sampling points. When applied as an NN, the multi-layer feedforward NN offers multi outputs. However, when used as DT's, the required number of DT's can be grown independently. Each DT is grown with the same input selection and the output is assigned to only one of the sampling points in the observation area. This is also applicable for NN's but it is not efficient to use a separate NN for each sampling point.

Full Decision Trees

The DT's are grown without pruning for different input selections. Afterwards, the eigenvalue region prediction is performed. The errors shown in table 3.8 are the sum of the false dismissal errors and the false alarms. Any eigenvalue outside the predicted region is counted as an error. In fact, some eigenvalues are outside the predicted region but still relatively close to it and therefore the predicted region can still be used for successful OSA. Thus, the errors are smaller for only those predicted regions which are totally unrelated to the positions of the eigenvalues.

Inputs	Selection	Learning	Testing
5	KM-05	1.65%	4.14%
10	KM-10	1.46%	5.03%
16	KM-16	0.99%	2.37%
20	KM-20	1.29%	5.03%
30	KM-30	1.36%	5.62%
40	KM-40	1.39%	4.73%
50	KM-50	1.32%	3.85%

Table 3.8: STD of differences between predicted and computed damping coefficients for regression tree growing under different input dimensions

Note that the selection by the GA shows the best results since the selection process by GA is optimized for DT-based eigenvalue region prediction. The mean and standard deviation of the damping difference $\Delta\zeta_{reg}$ and the frequency difference Δf_{reg} are shown in table 3.9 and table 3.10.

	Learning		Testing	
Inputs	Mean	**STD**	Mean	**STD**
5	1.197	0.820	1.168	0.798
10	1.190	0.819	1.127	0.768
16	1.197	0.813	1.159	0.797
20	1.196	0.808	1.133	0.782
30	1.196	0.813	1.160	0.817
40	1.193	0.813	1.161	0.787
50	1.200	0.814	1.167	0.776

Table 3.9: Mean and STD for $\Delta\zeta_{reg}$

	Learning		Testing	
Inputs	Mean	**STD**	Mean	**STD**
5	0.118	0.054	0.114	0.052
10	0.117	0.055	0.112	0.053
16	0.119	0.055	0.114	0.050
20	0.118	0.054	0.113	0.053
30	0.118	0.055	0.112	0.052
40	0.117	0.055	0.111	0.052
50	0.119	0.055	0.112	0.051

Table 3.10: Mean and STD for Δf_{reg} in Hz

3.11 Linear Regression Methods

In [45] a case study involving the Manitoba Hydro system is used to highlight the usefulness of multiple regression analysis in solving the oscillation source location problem. Multiple regression is a statistical tool used for analyzing relationships between a set of variables. Multiple regression refers to a regression model in which the fitted value of the response or dependent value (Y) is a function of the values of more than one predictor or independent (X) variables.

The multiple regression problem may involve linear or nonlinear fitting, such as the fitting of straight lines, polynomials of various orders, exponential forms, logarithmic forms etc. The most common form of multiple regression is multiple linear regression.

Multiple Linear Regression

Multiple linear regression is an extension of the simple linear model. In the simple linear regression model the response or dependent variable (Y_i) is assumed to be a linear function of a single independent or explanatory variable (X_i). Such that:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{3.28}$$

where the model shown in equation (5.1) is probabilistic since the error term ϵ_i is a random variable and β_0 and β_1 are unknown parameters of the model which have to be estimated based on available data:

$$(X_i, Y_i) = (x_i, y_i), (x_2, y_2), \ldots, (x_n, Y_n) \tag{3.29}$$

Multiple linear regression allows more than one independent or explanatory X variable. If there are k independent variables X_1,X_2,...,X_k then the model will appear as shown in the following:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_i X_{ki} + \epsilon_i \tag{3.30}$$

If regression parameters b_0,b_1,...,b_k are chosen as estimates of β_0,β_1,...,β_k respectively, then the regression estimate of the dependent variable Y based on independent variables x_i can be described by the relationship:

$$\hat{Y_i} = b_0 + b_1 X_{1i} + \cdots + b_i X_{ki} + \epsilon_i \tag{3.31}$$

This results in a residual or error, e, where:

$$e_i(error) = Y_i - \hat{Y_i} \tag{3.32}$$

Therefore, the best regression line would be the one in which the regression coefficients b_0 to b_i minimize the total errors or residual sum of squares (SSR):

$$SSR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{3.33}$$

In performing multiple linear regression analysis where a large number of independent variables or predictors may be available, the amount of independent variables to include in the final model becomes an important decision which has a profound effect on the model prediction capability and fit.

3.11.1 Variable Selection Methods

Due to the nature of the power system and the host of interactions that contribute to the damping, the corresponding regression model can get very complex. As with most models, a key objective is to keep it as simple as possible since the addition of redundant or insignificant parameters will only increase the error in the subsequent prediction. The other objective that partly stems from the first is not to exclude important variables which serve to reduce the error, as well as explaining some of the variance in the response variable.

There are a number of methods reported that use some sort of variable selection procedure to improve the statistical model. These normally involve some forward or backward eliminations steps to extract important variables.

The significance of a test is the probability the test statistic will reject the null hypothesis when the hypothesis is true i.e. the probability we observe our experimental results if we take samples from a population were there is no interacting effects [46]. After a coefficient estimation, the t-statistic of that coefficient can be calculated. This is the ratio of the coefficient to its standard error. At a given significance level, backward elimination starts with the largest possible model and looks at the individual t-statistics. At each iteration step the least significant variable gets dropped and the process continues until no more variables can be dropped.

A similar method is called forward selection where the process starts with the simplest model and adds variables as necessary. The first variable selected is the one that leads to the highest coefficient of determination (R^2) value. The R^2 value dictates the amount of variation explained in the dependent variable by the current independent variables. It is non-decreasing as more variables are added and eventually reaches 1 when the number of parameters equals the number of observations. If the variables contribution to the SSR is insignificant, then the variable is not included. This process continues until no significant variable to be entered can be found.

Stepwise regression is a compromise between the previous two methods. In both backward elimination and forward selection, once a variable gets removed it never gets entered again into the model. Stepwise regression starts with forward selection and includes the most significant variable X_j if its contribution $SSR(X_j|X_i)$ is significant and largest among all $X_j (j \neq i)$. Backward elimination is now used and the significance of $SSR(X_i|X_j)$ is checked. If it is not significant then it is dropped from the model and the next variable X_k for which $SSR(X_k|X_j)$ is significant and largest among the remaining variable is included. If $SSR(X_i|X_j)$ is significant, then both X_i and X_j are retained in the model and the search continues until no variable can be dropped [45].

Lastly, principle component analysis (PCA) may be used to reduce the number of variables in the model whilst retaining the important effects. There are a large number of loadflow variables in a typical power system and these are often strongly correlated. The objective of PCA is to take p variables X_1,X_2,...,X_p and find combinations of these to produce uncorrelated indices Z_1,Z_2,...,Z_p (principal components) that can be ordered in terms of their ability to describe the data. Where correlation exists between the original variables, it is likely that the data set can be adequately described using a subset of the principal components [2].

3.11.2 Development of a Linear Expression

If accurate parameters can be calculated, then a linear expression can be established which may be used to predict the damping of a mode based on measurable power system quantities, in real time. In the Manitoba case in, some of the 64 variables initially identified from the system were eliminated as their sensitivity to changes in damping were negligible over the range of power system operating points. This resulted in 37 variables that were interesting enough to be added to the regression model.

Note that, although a linear model is derived, most of the variables in this set of 37 variables are MW, MVAr and MVA quantities which are nonlinear functions of voltage magnitudes and phase angles. Therefore, in terms of voltage magnitudes and phase angles, the derived model in nonlinear.

The model was applied using stepwise regression. During the multiple linear regression analysis with stepwise selection, further variables were eliminated resulting in 12 independent variables in the final model.

Figure 3.7 shows a plot of the estimated damping, determined by multiple linear

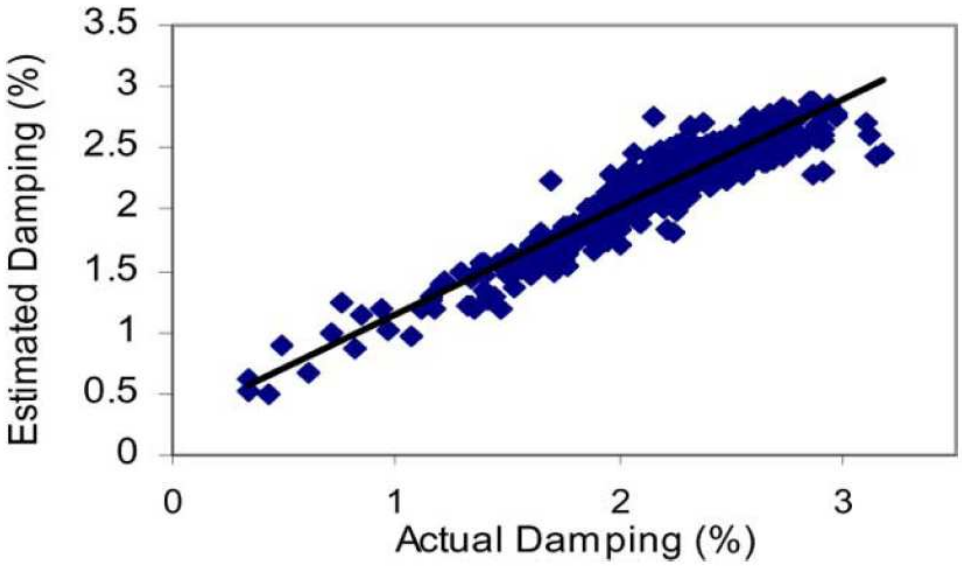


Figure 3.7: Estimated versus Actual Damping

regression versus actual damping with a trend line. The RMSE for these results was 0.11. This suggests that using this unique combination of 12 highly correlated vari-

ables, damping on the inter-area mode on average can be predicted to within ±0.11%. Parameters (β's) were calculated for the 12 independent measurable power system variables which could be used to predict the damping of the interarea mode.

Variable Name	Definition	Beta Value
X8	P_{gen} Bus 689	0.04115
X9	P_{flow} Bus 504/703	0.21069
X11	V_{mag} bus 506	-15.94898
X22	V_{ang} bus 510/503	0.00837
X31	P_{flow} Bus 703/504	0.36990
X32	P_{flow} Bus 703/509	0.14477
X35	P_{flow} Bus 506/507	0.12066
X37	MVA_{calc} Bus 506/507	0.07210
X2	V_{ang} bus 500	0.02166
X36	Q_{flow} bus 506/507	0.00193
X25	P_{flow} Bus 510/512	0.00470
X17	V_{ang} bus 508/503	0.02253

Table 3.11: β parameters for highly correlated variables

These variables are shown in table 3.11. The fact that the regression analysis identifies power flow from bus 506 to 507, which is the identical critical 230kV line that was highlighted in the observed trends section earlier in the literature, as an important variable among the final twelve selected, reinforces the practicality of the calculations. The results also suggest that to accurately predict the damping of the mode in question, the following additional information is needed:

- Voltage magnitude at the sending end bus (bus 506) of the critical 230kV line
- MW flow from the generator at bus 689
- MW flow from the two generators at bus 688 and 689 to the large mining load at bus 703
- Angular difference between the voltage phasors at bus 510 and 503

- Angular difference between the voltage phasors at bus 510 and 503
- Angular difference between the voltage phasors at bus 508 and 503
- MW flow from bus 504 to the large northern mining load at bus 703

If the required power system variables can be accurately measured at the identified locations and this data fed into the control room then the linear expression developed can be used to calculate damping of the dominant inter-area mode in real time. This may provide an additional accurate feedback tool for operators to gauge the dynamic state of the power system and aid in the decision making regarding generator dispatch, allowable transfer levels and system topology.
It is expected that prior to implementation, real-time data from the local WAMS will be used to benchmark and refine this tool.

3.11.3 Reliability of Regression Analysis

In this study around 755 system operating points were randomly determined. This resulted in 755 sets of damping ratios for the inter-area mode along with 755 sets of the independent, measurable power system variables. Further statistical analysis was performed to assess the minimum amount of data sets required to produce a reliable result with an acceptable RMSE. The last 105 sets were used to form an independent testing set and the first 650 used to form training sets beginning with a minimum set size of 50 and progressing in steps of 50 data sets to a maximum of 650. Figure 3.8 shows the results for a training set of size 650.

3.12 Summary

This chapter introduced some CI methods implemented for OSA. The errors shown depend not only on the quality of the prediction methods, but also on external factors such as data normalization and the distribution of the patterns used for training/learning and testing. The results may be influenced by the distribution. They also depend strongly on the selection and many other parameters inside the CI methodology. In NN training, the numbers of neurons and the training time are important issues.

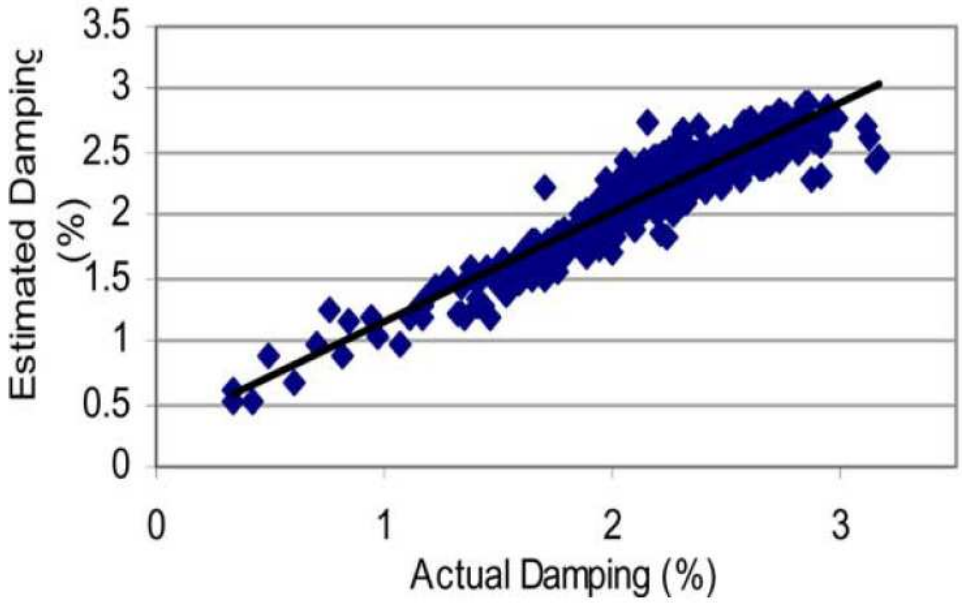


Figure 3.8: Estimated versus Actual Damping with full 650 Training Set

In general, the results of all introduced CI methods are accurate enough for online OSA. The most effective method is the decision trees, since it is the fastest in terms of data learning and gives the lowest errors over different input selections. The reason for the high performance is that each DT prevents the activation for only one sampling point, but the NN predicts a complete row of sampling points. When each sampling point is assessed by an individual NN, the NN results are improved. However, the method is not applicable to a large number of NN's due to the enormous time necessary for NN design and training. However, the DT method produced the most accurate eigenvalue region prediction, with training and testing sets having respective standard deviations between predicted and computed values of 0.99% and 2.37%.

The dependency of the number of inputs on the errors is much lower than one might expect. Even a selection of five inputs leads in most cases to tolerable results. But in terms of robustness, it is highly recommended to keep the number of inputs above five in order to create redundancy in the dataset.

In [45] it was seen how multiple regression analysis was used for estimating and pre-

dicting damping of a large interconnected power system. Based on previous system information from online recorders and offline studies the results confirm the practical usefulness of this approach. The Manitoba Hydro System model was used in this study with the state and dynamic data being used to develop a multiple regression model. The optimum results were achieved with a reduced variable model using 650 datapoints which achieved a RMSE of 0.11. This implies that 89% of the response variable is emplained by the predictors i.e. state variables such as active and reactive power.
The results indicate that previously observed trends in the power system could be useful in approximating the damping on the inter-area mode, but are not suffient to actually predict it. To accurately predict the damping of a mode, additional information which can be obtained by performing multiple regression analysis, is required. The information obtained could be useful in optimizing the location of online recording devices such as DSR's and PMU's.

If the required power system variables can be accurately measured at the identified locations and this data fed into the control room, then the resulting expression may be used to calculate damping of the dominant interarea mode in real time and to develop suitable rules for reacting to damping conditions.
Sensitivity analysis strengthens the case of this approach and may provide an additional feedback tool for operators to gauge the dynamic state of the power system. This would also aid in the decision making regarding generator dispatch, system loading, allowable transfer levels and system topology.

However, this approach has only been applied to data from the linear operation of the system where decay time constants are relatively small. During non-linear operation, these linear techniques with raw power system data cannot be used as the relationship between system state and damping is non-linear. In these cases, non-linear techniques and/or data transformation is required to deal with this non-linearity.

Chapter 4

Power System Modelling

4.1 Introduction

In conventional electrical power systems, the primary sources of electrical energy are the synchronous generators that must be kept in synchronism when interconnected to ensure small-signal stability at all times [4]. In addition, various control systems are used to ensure the system maintains synchronism and that oscillations in the system are controlled before they grow large enough to do permanent damage to the system. Understanding the underlying physical phenomena behind the various non-linear parts of the power system becomes more challenging for analysis and control design. Therefore, correct understanding and modeling of such components is vital for stability studies [1].

In this chapter, an overview of the test systems and how they were used to model the real system is presented. By carefully validating the source location methodology in terms of dynamic models, it is hoped that it can then be applied with confidence to the real power system.

4.2 Case Study Models

The testing presented in the following chapters consists of three case studies. Two are analytical dynamic models developed using matrix software and the other is a real system. Two interarea modes are used to test the source location method in the

dynamic models while data from a single troublesome mode in the real system is used to test the methodology on a real system.

The first dynamic model is a 16 machine five area test system that is used in the testing of the wavelet transformed general linear models. The second dynamic model is the Icelandic system that consists of 35 aggregated generators whose outputs are used to develop a statistical model with the mode decay time constant as the response variable.

The real power system consists of over 250 machines that can be used to develop a statistical model. In this case, mode information is extracted via PMU's at five different locations as a given mode may be more observable at different parts of the system, especially if they're physically close to the generators causing them.

4.3 Icelandic System

The transmission system in Iceland mainly operates at voltages from 66kV to 220kV, with the exception of a few 33kV lines. The transmission system is built up of a meshed 220kV network in the south part of the island connected to a 1000km long 132kV network that lies around the coast as shown in figure 4.1. Today a large part of the installed capacity of the system lies in the 220kV network, thus making it the stronger part of the system. In the 132kV ring network there is relatively less installed capacity making it the weaker part of the network.

Figure 4.1: Top level schematic of the Icelandic Power System

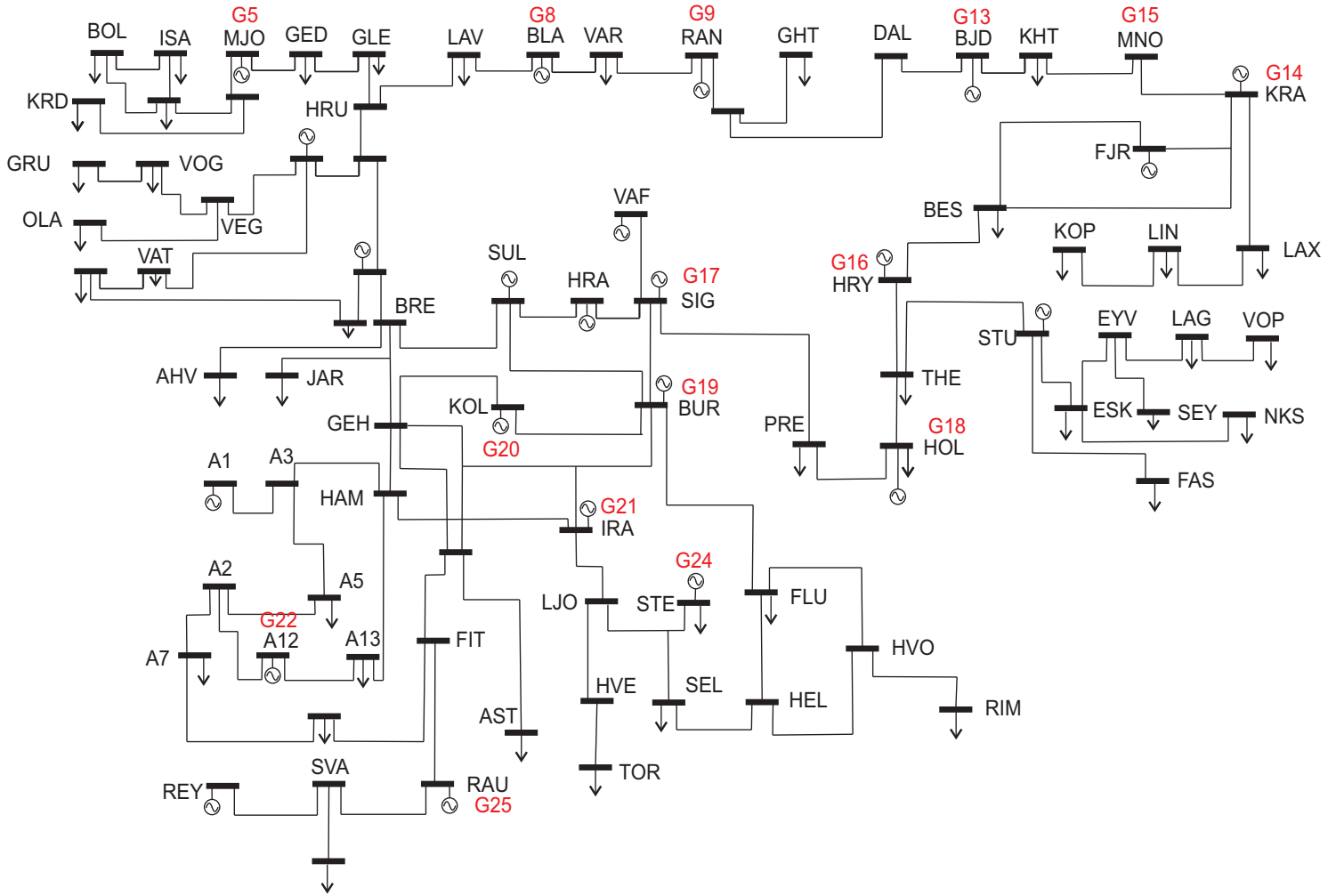


Figure 4.2: Single line diagram of the Icelandic Power System. Red generator numbers are shown along with bus names for ease of reference in later chapters where specific generators and groups of generators are highlighted as contributing factors to poor damping in the 0.21Hz and 0.62Hz Iceland models

A number of modes in the 132kV ring can be used to demonstrate various phenomena in the system. Common modes on the 132kV ring are those occurring around frequencies:

- 0.48Hz
- 0.60Hz
- 0.80Hz
- 0.95Hz
- 1.30Hz
- 2.00Hz

A study of the 0.21Hz interarea mode is presented later in this work as it involves the interaction of a number of generators across the system. It is clear that there is a pattern of poor damping at several of the above mode frequencies.

4.4 Dynamic Icelandic Model

The open-source MATLAB power system analysis toolbox (PSAT) was used to develop the system model. The dynamic model has a total of 35 fifth-order generators, 189 buses, 52 loads, 122 transformers (both two-way and three-way) and 84 transmission lines. The AVR's are enabled at all the generators with PSS's enabled at Blanda and Krafla to provide adequate damping of desired modes. Turbine governors are modelled on twelve of the generators. The transmission system voltages are mainly 220kV, 132kV, 66kV and 33kV. The model also contains a few buses with distribution voltages as low as 0.415kV. The generation voltages are 6.6kV, 11kV and 13.8kV. The system is shown in figure 4.1 and 4.2.

For stability studies and controller design, designers often adopt various simplified generator models that are recommended in the literature. Hence, for each synchronous generator a test model is represented by a fifth order dynamic model whose nonlinear differential and/or algebraic equations can be found in [4].

The loads in the Icelandic dynamic model are modelled as constant admittances. An advantage of this assumption is that it is possible to eliminate the network nodes to obtain an equivalent system which only consists of non-linear differential equations. In addition, it is possible to simplify the process of computing the new voltages at the network nodes by using sources of energy only [47]. In [48] it was shown that the type of load modeling used in the Greek system had little effect on the modal attributes of the system with combinations of constant admittance, constant power and constant current loads producing similar results.

Topological reduction is achieved by transforming the Icelandic network into a smaller equivalent network by either aggregation or elimination of nodes. This can be achieved as follows:

- Pre-fault load calculation is performed. The equivalent steady-state calculation is performed. The equivalent steady-state impedance loads are calculated in the form of admittances as:

$$y_{LK}^{*} = \frac{P_{LK} + jQ_{LK}}{V_k^2}, \quad k = 1, 2 \ldots \tag{4.1}$$

  for each load. Once these elements are calculated, they are added to the coresspond-ing matrix in the $\bar{Y}_{bus}$ admittance matrix.

- The internal machine voltages are calculated for each of the n machines in the network.

- $\bar{Y}_{bus}$ is augmented by admittances corresponding to internal machine nodes/buses and load buses. The admittance matrix can then be partitioned as:

$$\bar{Y}_{bus} = \begin{bmatrix} \bar{Y}_a & \bar{Y}_b \\ \bar{Y}_c & \bar{Y}_d \end{bmatrix}$$

  The nodal equations which represent the relation between the injected currents

and node voltages is given by:

$$\begin{bmatrix} \bar{I}_G \\ \bar{I}_L \end{bmatrix} = \begin{bmatrix} \bar{Y}_a & \bar{Y}_b \\ \bar{Y}_c & \bar{Y}_d \end{bmatrix} \begin{bmatrix} \bar{V}_G \\ \bar{V}_L \end{bmatrix}$$

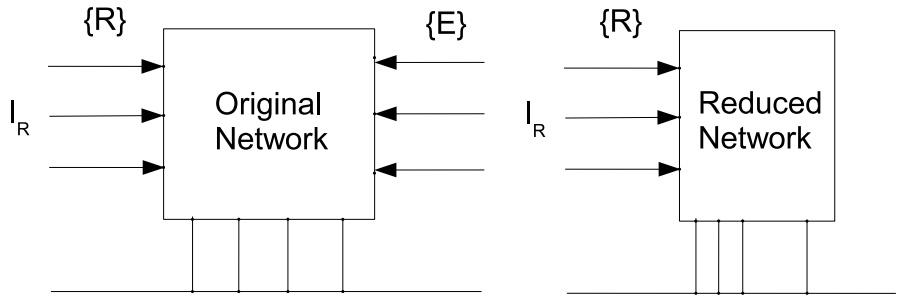
where $\bar{I}_G$ and $\bar{V}_G$ represent vectors of source currents and voltages respectively. $\bar{I}_L$ and $\bar{V}_L$ are vectors of currents injected by the loads and node voltages at load buses respectively. Since there are no injected currents in the network nodes, $\bar{I}_L = 0$, the system is reduced to internal machine nodes as follows:

$$\bar{V}_G = \frac{\bar{I}_G}{\bar{Y}_{int}} \tag{4.2}$$

where the matrix of admittance $\bar{Y}_i$ is given by:

$$\bar{Y}_i = (\bar{Y}_a - \bar{Y}_b \bar{Y}_d^{-1} \bar{Y}_b) \tag{4.3}$$

With the load nodes having zero current injection, the system matrix is reduced as shown above. Power systems are most readily described by non-linear differential



(a) Original network (b) Reduced network

Figure 4.3: Elimination of Nodes before and after elimination

and/or algebraic equations (DAE) which are a combination of dynamic models of individual components in the power system. The differential and algebraic equations governing the fifth order synchronous machines model used in the Icelandic model are given in [4] in p.u *dq* reference frame [49].

In p.u, the subtransient, transient and synchronous reactances are equivalent to the corresponding inductances. Therefore, common industry practice is to express synchronous machine parameters in terms of reactances instead of inductances. Figure 4.4 summarizes the overall modeling and simulation of a typical power system.

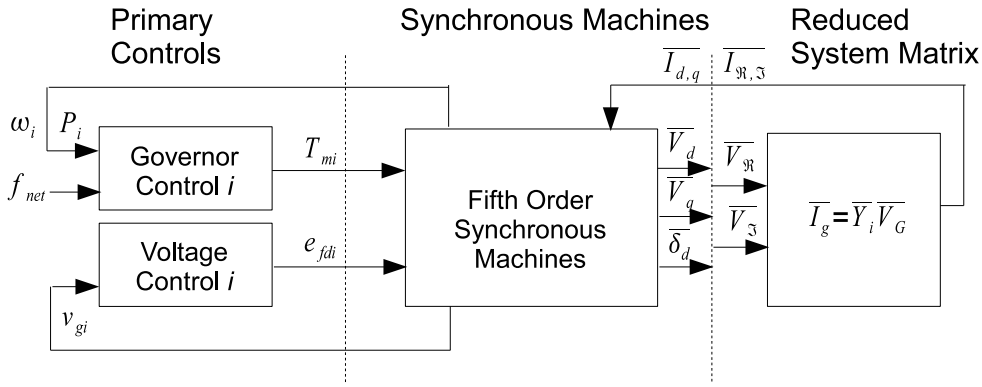


Figure 4.4: Simulation of power system with constant admittance load type

4.4.1 Small Signal Analysis

PSAT uses a full Newton-Raphson algorithm with a slack bus for solving the power flow problem. The Jacobian matrix is updated every iteration until the set tolerance is reached and the dynamic model, linearized about the steady-state operating point, is obtained. Using a separate algorithm, the eigenvalues can be obtained from the state matrix to give both frequency and damping values. Phase and amplitude information is extracted from the output matrix. The model produces a total of 373 eigenvalues together with 373 state variables.

The left and right eigenvectors and the derived participation factors may also be used to identify which modes are rotor angle modes and which generators participate in them. Additionally, active load modulation may be examined using input matrices representing the effect of loads on the system [50].

The eigenvalues for a given mode are tabulated over all the loadflow scenarios that make up an eigenvalue time domain simulation of the system. It is the decay time constants and damping ratios derived from the state matrix that are used to build

successive statistical models. The eigenvalue plots for a given simulation can be seen in figure 4.5.

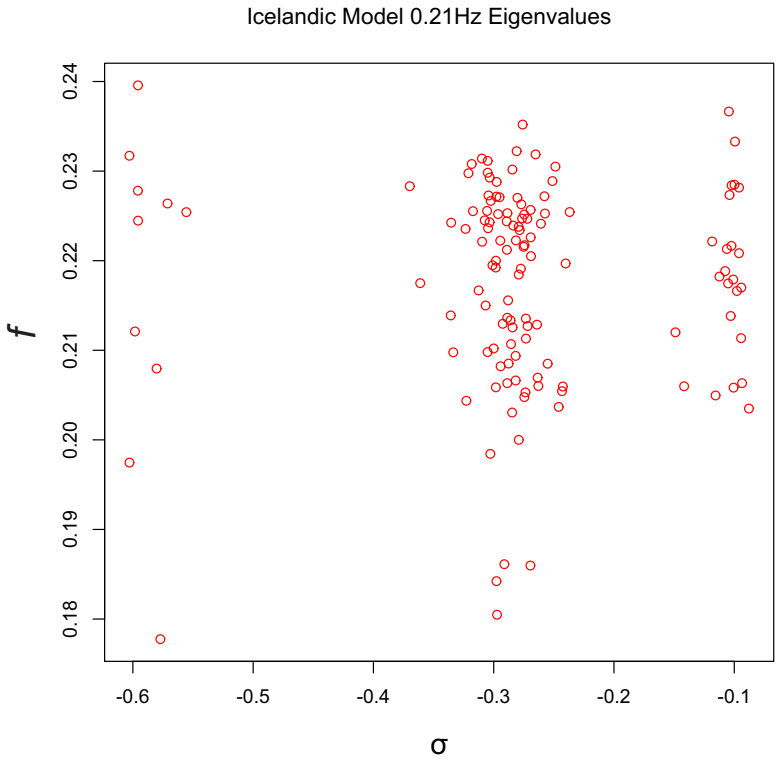


Figure 4.5: Eigenvalues of 0.21Hz interarea mode from the Icelandic Model with 128 data points

4.5 Sixteen Machine Five Area Test System

This dynamic model is a reduced order equivalent of the interconnected New England test system (NETS) and the New York power system (NYPS) which can be found in [52]. The test system contains 68 buses out of which sixteen buses are generator buses as shown in figure 4.6. There are five geographical regions out of which NETS and NYPS are represented by a group of synchronous generators. The remaining areas 3,4 and 5 are approximated by equivalent generator models. The power exchange between NETS and NYPS is transferred through major transmission corridors by the connecting bus numbers 60-61, 53-54 and 27-53. All these transmission corridors have double circuit tie lines which carry total tie line power exchange between NETS and NYPS of approximately 700MW.

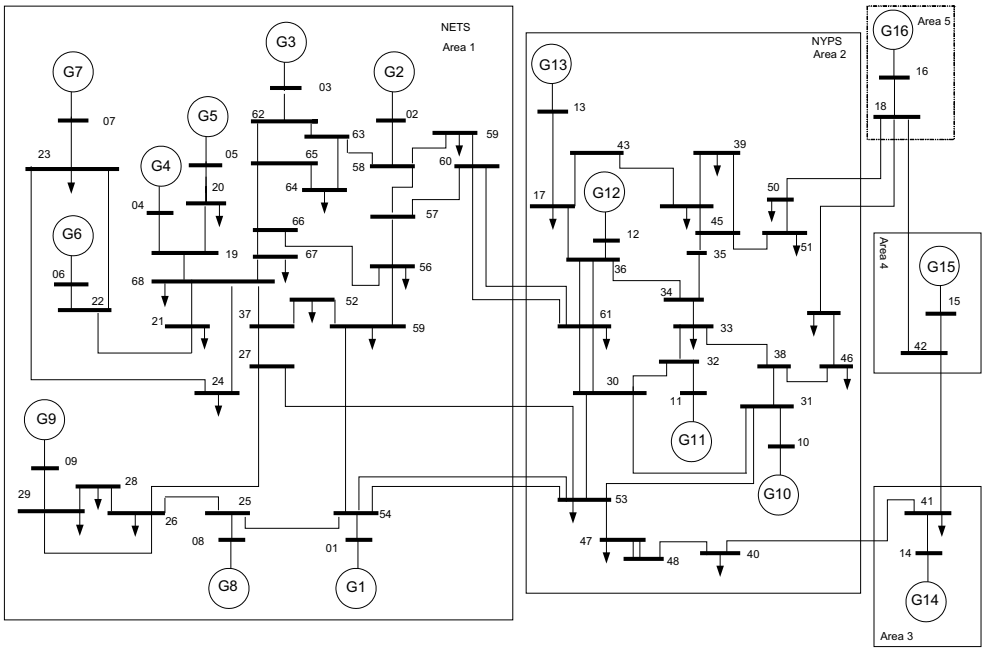


Figure 4.6: Model Schematic of the New England test system (NETS) and the New York power system (NYPS)

All the synchronous generators are represented by fifth order dynamic models similar to the other test systems. G1 to G8 are equipped with IEEE slow excitation systems, namely IEEE-DC1A models. However, G9 is equipped with a fast acting static excitation system, IEEE-AC4A and a speed based power system stabilizer (PSS) to ensure adequate damping of the local oscillations [4] [52]. This is the only machine equipped with a local PSS.

G14 and G16 are also equipped with high-initial IEEE static excitation systems, IEEE-AC4A. The rest of the generators is under manual excitation control. The reduction of the system matrix is achieved by matrix algebra as shown previously, on the assumption that all loads are treated as constant admittance i.e. zero current injection.

4.6 Generating Multiple Power Flow cases for Eigenvalue analysis

When there are distinct changes in the dynamic performance of the network, changes in the system state must be considered. The load flow data must be analyzed to determine if any significant system condition changes correspond to a change in damping. In some networks, many changes may take place simultaneously and it may be hard to determine which states are most associated with dynamic performance. The test systems used in this study were the Icelandic power system and the sixteen machine five area system that can be found in [1] [50].

4.6.1 Random Variable Adjustment

The starting point for generating a large number of powerflow cases is a solved base case powerflow which contains the base values around which adjustments may be made to simulate the real power system. The following model is used to randomly generate power flow cases at the various system operating points whilst not violating the preset limits that maintain realistic operation of the system. Examples of generator tolerances can be found in [53].

For the kth operating point, active and reactive power consumption at the ith load bus is given by [45]:

$$P_L^{(k)} = P_{LO}^{(i)} \left\{ 1 + 2\Delta P^{(i)} \left[0.5 - \epsilon_{PL}^{(i)}(k) \right] \right\} \tag{4.4}$$

$$Q_L^{(k)} = Q_{LO}^{(i)} \left\{ 1 + 2\Delta Q^{(i)} \left[0.5 - \epsilon_{QL}^{(i)}(k) \right] \right\} \tag{4.5}$$

where ${P_{LO}}^{(i)}$ and ${Q_{LO}}^{(i)}$ are real and reactive power consumption at the ith load bus in the base case; $\Delta P(i)$, $\Delta Q(i)$ are allowable fractional adjustments of real and reactive power at the ith load bus; and ${\epsilon_{PL}}^{(i)}$, ${\epsilon_{QL}}^{(i)}$ are uniform independent random variables between 0 and 1.

Similarly, the active power injections and generator voltage setpoints on a given generator regulating bus (k) can be randomly generated about the values in the base case

and within the prescribed operating limits as given by:

$$P_G^{(k)} = P_{GO}^{(i)} \left\{ 1 + 2\Delta P^{(i)} \left[0.5 - \epsilon_{PG}^{(i)}(k) \right] \right\} \tag{4.6}$$

$$V_G^{(k)} = V_{GO}^{(i)} \left\{ 1 + 2\Delta V^{(i)} \left[0.5 - \epsilon_{VG}^{(i)}(k) \right] \right\} \tag{4.7}$$

where ${P_{GO}}^{(i)}$ and ${V_{GO}}^{(i)}$ are base case active power injections and voltage setpoint settings at the ith generator bus; $\Delta P(i)$, $\Delta V(i)$ are allowable fractional adjustments of active power injection and generator bus voltage setpoint at the ith generator bus, and ${\epsilon_{PG}}^{(i)}$, ${\epsilon_{VG}}^{(i)}$ are uniform independent random variables between 0 and 1.

The parameters ${P_{LO}}^{(i)}$, ${Q_{LO}}^{(i)}$, ${P_{GO}}^{(i)}$ and ${V_{GO}}^{(i)}$ are the critical parameters in the random variable adjustment. As intentional variable adjustment is used later, the random variables were set to vary by a relatively small amount to simulate the continual activity of the system. $P^{(i)}$ and $Q^{(i)}$ were set at 0.2 so that the active and reactive generation at selected buses could vary between 0.8 and 1.2 times the nominal value. $V^{(i)}$ was set at 0.1 which represented the maximum voltage deviation of selected buses which meant that nominal voltage could only deviate by a factor of 0.1.

For each adjustment of the parameters the loadflow was solved and an eigenvalue analysis was performed. The active and reactive power generation as well as the line flows and bus voltages were recorded. The amplitude, decay time constant and frequency of the system modes were also recorded for use later in the statistical modeling.

4.6.2 Intentional Variable Adjustment

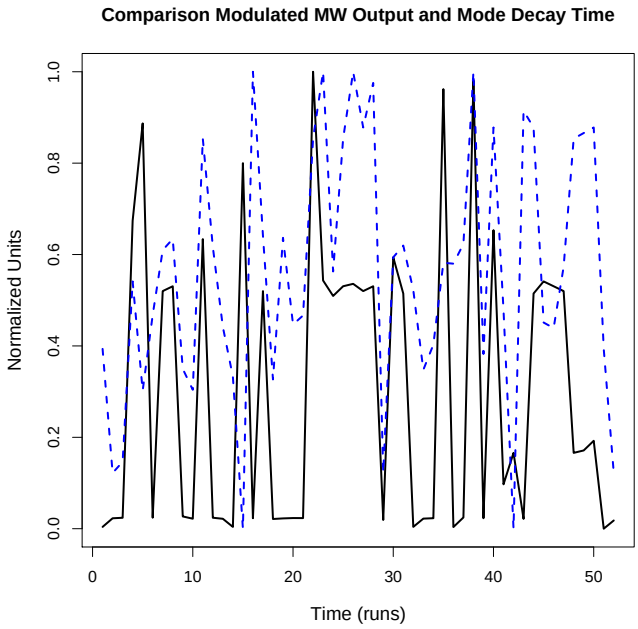
The random variable adjustment detailed earlier is useful in creating system conditions in a model. However, the preset limits of the synchronous machines frequently prevent adequate system conditions from being realized that create interesting dynamics and lengthy decay time constants. In the case of both the sixteen machine and Icelandic systems, the random variable adjustment didn't provide the necessary dynamic behaviour as often as required due to the relatively small changes implemented. In addition, for testing purposes the reproduction of each decay time constant profile is

desirable so that the profiles before and after the generator modulation can be compared. Modulation refers to intentionally changing generator output in an attempt to increase or decrease mode decay time constants.

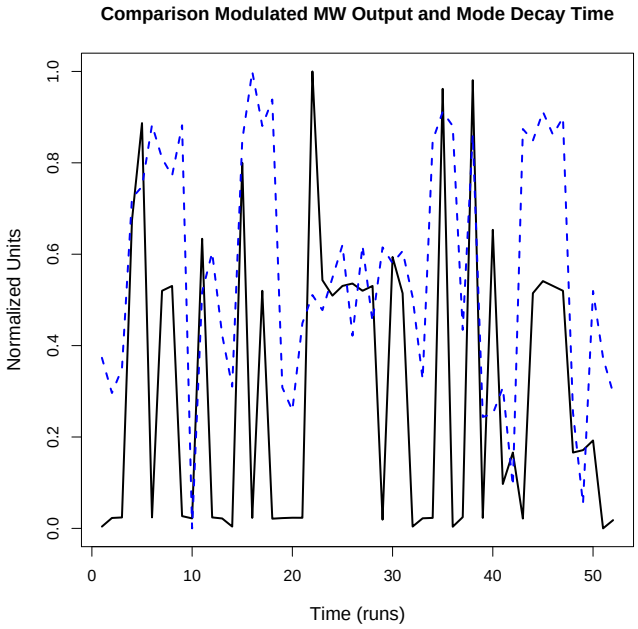
By implementing the random variable adjustment as well as modulating individual generators it is possible to produce the desired decay time constants as well as rendering the simulations reproducible. This is achieved by increasing certain generators to their limits (or to a fraction of their limits) at various points along a given window which allows some of the interesting dynamics to be realized. Usually, the systematic increases in generator outputs are much larger than the random injections as generators are pushed closer to their limits. In some instances, generators can be ramped from 0.1 - 1.8 times their nominal value which in turn should provide a shift in the certain system dynamics as long as the generator(s) in question are participating.
Of course, some trial and error is required in order to determine the correct combinations of machines that, when set at their limits, produce the long decay time constants required.
In addition, changing known groups of generators makes the testing procedure easier as it is possible to compare results from the source location methodology to the known generators that cause the oscillations in the first place. In essence, the model is run with underlying random variable adjustment (simulating small kicks in the system) with larger intentional generator output changes superimposed to produce resonant modes.

However, the main advantage of this setup is the reproducibility of the decay time constant profiles due to the fact that the random variable adjustment only causes relatively small fluctuations in the decay time constant whereas the intentional changes in the selected machines cause much larger fluctuations. This means that re-running the simulations with the same intentional changes, allows the same decay time constant profile to be produced each time with only relatively small differences in the random fluctuations.

Figure 4.7 shows an example of the modulation process at work. In this particular



(a) Intentional Modulation G1



(b) Intentional Modulation G2

Figure 4.7: Plots showing significant generator outputs (blue) against mode decay time constant

example ten generators were modulated along the entire simulation period. It must be noted that they weren't all changed at once, rather groups of them were modulated at the beginning of the window and other groups modulated at various times as the simulation progressed. The subfigures show two of the generators from the Icelandic model that have been adjusted for a period of time over the window.
They have been set at their maximum values at different points across the window. At various instances when the generators are high, the decay time constants also seem to be quite high. This may be coincidental, however, by modulating a number of machines in this way it is possible create poorly damped conditions as well as distinctive generation patterns in the system for comparison. The generators responsible may be extracted using linear techniques or logic regression on the data derived from these poorly damped periods.

The main benefit of this approach is that the reported generators can be tested quite easily with the model. This is simply done by replacing a given generator with a low constant power source (to maintain system topology). By modulating the same generators as before (except the generator(s) is question) and by implementing the underlying random injections, a similar decay time constant profile can be created without the effect of the reported generator. By comparing the decay time constant profiles of a number of simulation runs with the constant power source against the decay time constant profile of the full model it is possible to determine the effect the removed generator (constant power source) has on the mode damping.

The constant power source is implemented by introducing a for loop into the loadflow runs with the effect of setting a desired generator to the same power for every loadflow run i.e. holding it constant. This testing process is applied to all the dynamic models in the subsequent chapters of this thesis as a means of visualizing the effect a given generator or group of generators has on mode damping. By doing this, it is possible to validate the results from the source location methodology.

4.7 Summary

In power systems, synchronous generating units are the main source of electrical energy supply via high-voltage transmission lines and low voltage distribution systems. The stability of the power system depends on the controllers connected to the system such as excitation systems. Therefore, proper understanding and modeling of characteristics of such components is a matter of fundamental importance for stability studies.

In this chapter, the aim was to present the readers with proper understanding of the processes that are used to model the real power system. Constant admittance loads were discussed as a means to reduce the network size and make it more computationally friendly. The use of multiple loadflow scenarios along with random and systematic or intentional injection models were presented as viable methods of producing interesting dynamics in the system models. Random injections were limited to $\pm 20\%$ nominal injection to simulate constant system activity whereas the systematic intentional injections pushed the generators to their limits.

The tabulated values of the decay time constants, as well as the active and reactive power injections are used later on to develop statistical models for use in source location. In the following chapter both the sixteen machine and Icelandic model will be used to demonstrate the usefulness of the source location technique. This work will culminate with the validated methodology being applied to real data from a real system which exhibits some lengthy decay time constants associated with complex dynamics.

Chapter 5

Investigating Events using Wavelets

5.1 Introduction

Interest in wavelet analysis has been growing very rapidly in recent years. Advances in its theory, algorithms and applications have greatly influenced the development of many disciplines of science and technology including mathematics, physics, engineering, geosciences and meteorology.

The theory of wavelets stands at the frontiers of mathematics, scientific computing and signal processing. Its goal is to provide a coherent set of concepts, methods and algorithms that are adapted to a variety of non-stationary signals and which may also be used as a tool for signal processing [54].

Wavelets and wavelet packets constitute useful tools for the decomposition of complicated functions into a small number of elementary waveforms that are localized both in time and frequency [55] [56]. The main ideas of their construction can be found in [89]. They have an advantage over traditional Fourier methods in analyzing physical situations where the signal contains sharp discontinuities or spikes.

The fundamental idea behind wavelets is to analyze according to scale. Approximation using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superimpose sines and cosines to represent other functions.

However, in wavelet analysis, the scale that we use to look at data plays a special role. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large "window" we would notice gross features. Similarly, if we look at a signal with a small "window," we would notice small features. The result in wavelet analysis is to see both scales as well as those that lie inbetween.

This makes wavelets interesting and useful. For many decades, scientists have wanted more appropriate functions than the sines and cosines which comprise the bases of Fourier analysis, to approximate choppy signals [58]. By their definition, these functions are non-local and stretch out to infinity. As a result, they do a very poor job in approximating sharp spikes. However, with wavelet analysis we can use approximating functions that are contained neatly in finite domains which makes wavelets well-suited for approximating data with sharp discontinuities.

The wavelet analysis procedure requires a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a combination of the wavelet functions), data operations can be performed using the corresponding wavelet coefficients. If the best wavelets are chosen for the data, or the wavelet coefficients are truncated below a threshold, the data can be sparsely represented. This sparse coding makes wavelets an excellent tool for data compression [59].

A detailed description of wavelets as well as some of the variable selection methods used to improve the statistical models are presented in Appendix A. The appendix contains mathematical descriptions of the methods used as well as some visual examples and it may be useful for the reader to become familiar with Appendix A before embarking on the remainder of this chapter.
The rest of this chapter deals with the use of wavelet transformed variables in general linear models. Various reduction techniques are also applied to the data to improve

model fits and these are explored in depth.

5.2 Wavelet Transform vs. Fourier Transform

It is well known from Fourier theory that a signal can be expressed as the sum of a series of sines and cosines. This sum is also referred to as a Fourier expansion. The big disadvantage of a Fourier expansion however is that it has only frequency resolution and no time resolution. This means that although we might be able to determine all the frequencies present in a signal, we do not know when they are present. To overcome this problem, several solutions have been developed which are able to represent a signal in the time and frequency domain at the same time.

The idea behind these time-frequency joint representations is to cut the signal of interest into several parts and then analyze the parts separately. It is clear that analyzing a signal this way will give more information about the location and instance of different frequency components. However, it also leads to a more fundamental problem as to where to cut the signal. If the requirement is to know all the frequency components present at a certain moment in time, then only a very short time window is cut out using a Dirac pulse. It can then be transformed into the frequency domain to determine the frequency content of a signal at a given time. However, this procedure actually leads to an unacceptable results [60].

The problem here is that cutting the signal corresponds to a convolution between the signal and the cutting window. Since convolution in the time domain is identical to multiplication in the frequency domain and because the Fourier transform of a Dirac pulse contains all possible frequencies, the frequency components of the signal will be smeared all over the frequency axis. It is important to note that we are talking about a two-dimensional time-frequency transform and not a one-dimensional transform. In fact this situation is the opposite of the standard Fourier transform since time resolution is available but not frequency resolution.

The underlying principle of the phenomena just described is due to Heisenberg's uncertainty principle, which, in signal processing terms, states that it is impossible to know the exact frequency and the exact time of occurrence of this frequency in a signal. In other words, a signal cannot be simply represented as a point in the time-frequency space. The uncertainty principle shows that how the signal is cut is very important.

The wavelet transform is the most recent solution to overcoming the shortcomings of the Fourier transform. In wavelet analysis, the use of a fully scalable modulated window solves the signal-cutting problem. The window is shifted along the signal and for every position the spectrum is calculated. Then this process is repeated many times with a slightly shorter (or longer) window for every new cycle. In the end, the result will be a collection of time-frequency representations of the signal, all with different resolutions. Because of this collection of representations the wavelet transform may be described as multiresolution analysis. Wavelets are also referred to as time-scale representations, with scale representing the various window lengths over which the signal is analyzed [61].

In this chapter, wavelets serve a number of purposes in relation to the source location method. Although not explicitly used in this work, wavelets can be used to trigger alarms for poorly damped events. This is due to the fact that large wavelet coefficients at a given frequency represent large changes in a signal at that frequency. Therefore, a large, high frequency change in the signal will be represented by a large wavelet coefficient at high resolution.

Wavelets may also be used to linearize the relationship between mode decay time constants/damping ratio and the system state variables as will be shown later in this chapter. This allows linear techniques to be used on the transformed data as opposed to raw data which is non-linear close to the stability limits.

An additional use of the wavelet transformed variables is in the removal of redundant information from the dynamic and system state signals used to develop the statistical models. This attribute stems from the fact that the wavelet transform allows the

signals to be deconstructed into various frequency components. By determining low variance, low mean frequency components it is possible to remove this redundant data with the aim of improving the overall model.

5.3 Statistical Models for Source Location

In the previous section we discussed the advantages of using wavelets for event detection and removing redundant frequencies in the signal. It was also suggested that wavelets could be used to improve statistical models to produce better fits and thus better predictions. In terms of source location, the prediction is not of direct importance as it is the predictor regression coefficients and their significance that highlight important predictors that can be used to control damping.

However, the relationship between the predictors (generator active power, line flows etc.) and the response variables (decay time constants, damping ratio) is a non-linear one close to the stability limits. As well as this, the response variable error (damping or decay time constant) distribution is not normally distributed. If these conditions are not met then it is unwise to proceed with a linear regression analysis on the raw data.
Fortunately, there are a number of ways around this problem and a few of them are discussed here along with some of the assumptions that are implicit in the source location problem.

5.3.1 Linear Regression

It has been proposed that the sensitivities of oscillatory modes to static parameters can yield information on the operational variables that most strongly influence particular dynamic conditions [2]. Multiple linear regression is a statistical tool used for analyzing relationships between a set of variables. A detailed explanation is beyond the scope of this thesis and additional information can be found in the literature [62] [63].

Multiple regression analysis refers to a regression model in which the fitted value of the

response or the dependent variable (Y), is a function of the values of more than one predictor or independent (X) variable. The multiple regression problem may involve linear or non-linear fitting, such as the fitting of straight lines, polynomials of various orders, exponential forms, logarithmic forms etc. The most common form of regression is multiple linear regression.

Multiple Linear Regression

Multiple linear regression is an extension of the simple linear model. In the simple linear regression model the response or dependent variable (Y_i) is assumed to be a linear function of a single independent or explanatory variable (X_i). Such that:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{5.1}$$

where the model shown in equation (5.1) is probabilistic since the error term ϵ_i is a random variable and β_0 and β_1 are unknown parameters of the model which have to be estimated based on available data:

$$(X_i, Y_i) = (x_i, y_i), (x_2, y_2), \ldots, (x_n, Y_n) \tag{5.2}$$

Multiple linear regression allows more than one independent or explanatory X variable. If there are k independent variables X_1,X_2,...,X_k then the model will appear as shown in the following equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_i X_{ki} + \epsilon_i \tag{5.3}$$

If regression parameters b_0,b_1,...,b_k are chosen as estimates of β_0,β_1,...,β_k respectively, then the regression estimate of the dependent variable Y based on independent variables x_i can be described by the relationship:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + \cdots + b_i X_{ki} + \epsilon_i \tag{5.4}$$

This results in a residual or error, e, where:

$$e_i(error) = Y_i - \hat{Y_i} \tag{5.5}$$

Therefore, the best regression line would be the one in which the regression coefficients b_0 to b_i minimize the total errors or residual sum of squares (SSR):

$$SSR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y_i})^2 \tag{5.6}$$

In performing multiple linear regression analysis where a large number of independent variables or predictors may be available, the amount of independent variables to include in the final model becomes an important decision which has a profound effect on the model prediction capability and fit.

5.3.2 Variable Selection Methods

Due to the nature of the power system and the host of interactions that contribute to the damping state, the corresponding regression model can get very complex. As with most models, a key objective is to keep it as simple as possible since the addition of redundant or insignificant parameters will only increase the error in the subsequent prediction. The other objective that partly stems from the first is not to exclude important variables which serve to reduce the error as well explain a portion of the variance in the response variable.

There are a number of methods reported that use some sort of variable selection procedure to improve the statistical model. These normally involve some forward or backward eliminations steps to extract important variables.

In statistical terms, the significance of a test is the probability that the result didn't occur by chance [46] i.e. that the response variable behaviour was genuinely caused by some definite changes in the predictor variables. After a coefficient estimation, the t-statistic of that coefficient can be calculated. This is the ratio of the coefficient to its standard error. At a given significance level, backward elimination starts with the

largest possible model and looks at the individual t-statistics. At each iteration step the least significant variable gets dropped and the process continues until no more variables can be dropped.

A similar method is called forward selection where the process starts with the simplest model and adds variables as necessary. The first variable selected is the one that leads to the highest coefficient of determination (R^2) value. The R^2 value dictates the amount of variation explained in the dependent variable by the current independent variables. It is non-decreasing as more variables are added and eventually reaches 1 when the number of parameters equals the number of observations. If the variables contribution to the SSR is insignificant, then the variable is not included. This process continues until no significant variable to be entered can be found.

Stepwise regression is a compromise between the previous two methods. In both backward elimination and forward selection, once a variable gets removed it never gets entered again into the model. Stepwise regression starts with forward selection and includes the most significant variable X_j if its contribution $SSR(X_j|X_i)$ is significant and largest among all $X_j (j \neq i)$. Backward elimination is now used and the significance of $SSR(X_i|X_j)$ is checked. If it is not significant then it is dropped from the model and the next variable X_k for which $SSR(X_k|X_j)$ is significant and largest among the remaining variables is included. If $SSR(X_i|X_j)$ is significant, then both X_i and X_j are retained in the model and the search continues until no more variables can be dropped.

Lastly, principle component analysis (PCA) may be used to reduce the number of variables in the model whilst retaining the important effects. There are a large number of loadflow variables in a typical power system and these are often strongly correlated. The objective of PCA is to take p variables X_1,X_2,...,X_p and find combinations of these to produce uncorrelated indices Z_1,Z_2,...,Z_p (principal components) that can be ordered in terms of their ability to describe the data. Where correlation exists between the original variables, it is likely that the data set can be adequately described using a subset of the principal components [2].

Multicollinearity

Power system data may often comprise more independent variables than observations, especially in the case of high resolution events i.e. events were correlation may only take place over a short window. This case is encountered less in other applications of statistics. Collinearity of the independent variables is typical for loadflow data, that is, certain independent variables can be practically represented as a linear combination of other independent ones. This is the source of many problems in direct application of many statistical methods, such as the multiple linear regression (MLR).
Studies have shown that if collinearity is present independent among variables, the subsequent prediction capability can be inhibited. This has prompted the development of other linear methods to highlight redundancy in the independent variables. Several alternatives that are able to adapt to this multicollinearity were covered in previous sections which include stepwise regression and principal component analysis.

In power engineering, a number of linear statistical methods have been applied to solve quantitative problems with the assumption that the relationship between the dependent and independent variables is linear. However, it has been shown that this assumption leads to unacceptable loss of information and thus unacceptable results from the statistical model when the system approaches instability [2]. In the next section, partial least squares regression (PLSR) is introduced as a means to combat the problem of multicollinearity.

As well as PLSR, mutual information (MI) will be presented in the next section as a method to remove redundancy from the model. A key step in the source location algorithm is the removal of the multicollinearity and any redundancy which can be achieved with a combination of PLSR and MI. By determining the suitable variables to put forward into a model we can ensure that most of the information will be retained whilst the model fit is improved [64]. Figure 5.1 shows the usefulness of the PLSR method in removing multicollinearity from the regression data. Here, the PLSR step was applied to the Icelandic 0.21Hz mode that was used as a test system for each step

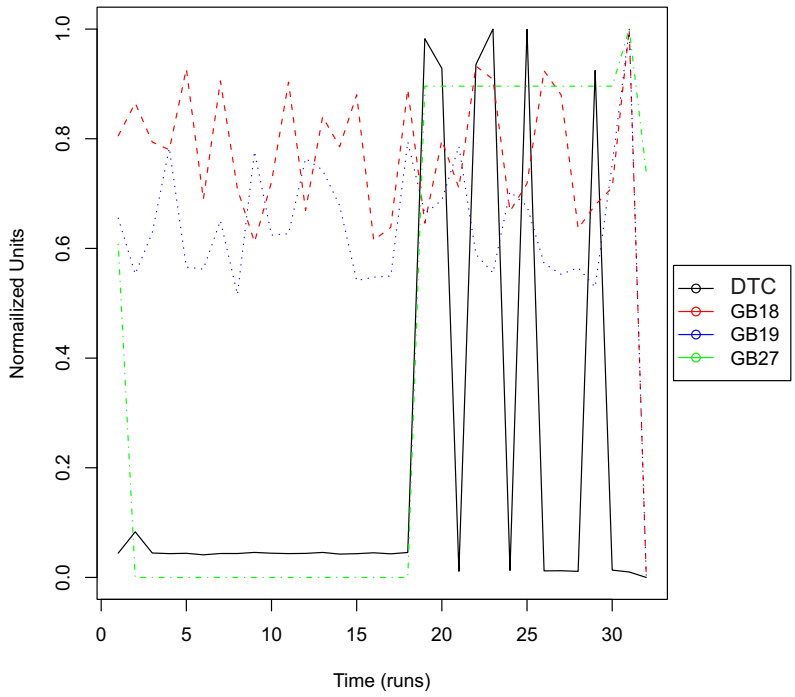


Figure 5.1: Highlighting multicollinearity in G6(GB18), G7(GB19) and G8(GB27) in Icelandic model using PLSR. DTC stands for decay time constant

of the methodology. In this case G7, connected to generator bus 19 or GB19, is singled out for removal as it is deemed to be less significant than its neighbour G6 which is connected to bus 18. The removal is achieved by identifying the collinear variables and removing the independent variable with the lowest single correlation in relation to the dependent variable. The collinearity can be clearly seen in figure 5.1 as both active power outputs of G7(GB18) and G8(GB19) seem to track each other for most of the window. In this case, retaining both variables will cause larger standard errors to be attributed to both GB18 and GB19. By determing the most important variable, the prediction of the model as a whole is maintained and the standard error of the remaining variable is reduced thus increasing its t-score or statistical significance.

Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a technique that generalizes and combines features from principal component analysis and multiple regression. It is particularly useful when the aim is to predict a set of dependent variables from a large set of independent variables. The goal of PLSR is to predict a dependent variable ($\mathbf{Y}$) from

a set of observations ($\mathbf{X}$) and to describe their common structure. When $\mathbf{Y}$ is a vector and $\mathbf{X}$ is full rank, this goal could be accomplished using *ordinary multiple regression*. However, when the number of predictors is large compared to the number of observations, $\mathbf{X}$ is likely to be singular and the regression approach is no longer feasible due to *multicollinearity*. Some approaches to this problem have already been discussed such as the use of stepwise elimination and principal component analysis (PCA).

In contrast to PCA, PLSR finds components from $\mathbf{X}$ that are also relevant to $\mathbf{Y}$. It also maintains the orthogonality of its components which means it is well suited to removing multicollinearity. Specifically, PLSR searches for a set of components, called *latent vectors*, that perform a simultaneous decomposition of $\mathbf{X}$ and $\mathbf{Y}$ with the constraint that these components explain as much of the covariance as possible between $\mathbf{X}$ and $\mathbf{Y}$. This step generalizes PCA and is followed by a regression step where the decomposition of $\mathbf{X}$ is used to predict $\mathbf{Y}$ [65]. Partial least squares regression produces factor scores as linear combinations of the original predictor variables, so that there is no correlation between the factor score variables used in the predictive regression model. Table 5.1 details a summary of the partial regression step applied to the Icelandic

Variable No.	Proportional Var	Normalized Values
1	0.033	0.442
7	0.042	0.563
13	0.046	0.621
18	0.052	0.702
25	0.074	1.000
27	0.064	0.864

Table 5.1: Summary of PLSR Regression results showing explained variance for the Icelandic Model

model (see section 5.9). After initial thresholding the model began with 30 variables i.e. four variables were removed. In an attempt to extract useful variables, 27 cross-validation steps were performed with different combinations of predictors. PLSR is a linear technique so it was performed on the wavelet transformed data as opposed to the raw dataset.

PLSR works by trying to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. In this work cross-validation between the model predictors was used. The model was iterated after each cross-validation to determine the performance of different combinations of predictors that make up each factor. As a result different *latent vectors* consisting of *principal components* were compared to see how well they explained the variance in the Y space i.e. the response variable.

In table 5.1 it can be seen that PLSR with 25 components (predictors) has the highest regression coefficient and best explains the Y space variance as it has the highest proportional variance i.e. explains the largest proportion of the Y-Space variance. The top scoring factors can be compared in terms of their components (original predictors) and the variables shown to be absent from the top vectors can be removed. As a result, original collinear components removed from the latent vectors (to maintain orthogonality) are removed with the most significant collinear variable being retained. As a result, a total of five predictors can be removed from the model.

The remaining components correspond to generator outputs that can be used to build an accurate statistical model. Of the variables removed, some may be collinear with others and may have been removed as they do not adequately account for Y-space variance in conjunction with other variables in the latent vectors. This has the effect of immediately reducing the standard error of the remaining collinear variable(s) which are more significant. The removed collinear variables may be re-introduced post-analysis to control mode damping i.e. collinear generators sitting in the same plant that were removed earlier.

Mutual Information

Most methods of variable selection use linear transformations to decide which variables (x) are most strongly related to the output data (y) being modelled. However, these methods can only pick up predominantly linear relationships and will tend to miss variables which have a strong but non-linear relationship with the output. If a

linear modeling technique is being used, then this can only be comprehensive when the relationships in the system are predominantly linear. Consequently, it is ill advised to select variables using a method, which prefers linearity, on data that is not linear. Clearly, a method which does not impose the criterion of linearity would be advantageous. This is where *mutual information* can be used on data that is not linearally related or nonparametric.

Mutual information [66] can be regarded as a generalized version of correlation. Where correlation assumes linear relationships and normally distributed data, mutual information makes no assumptions about the two data series being compared. The mutual information between a class c and an input feature f (with N_f components) is the amount to which the knowledge provided to the feature vector decreases the uncertainty about the class. Intuitively, mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other.
Mutual information is derived by calculating the probability distributions of the two series, $p(x)$,$p(y)$ and $p(x, y)$. It then compares the joint probability $p(x, y)$ with $p(x)p(y)$. For statistically independent data [67]:

$$p(x)p(y) = p(x, y) \tag{5.7}$$

Hence if these quantities are not the same, a dependence exists between the two data series and this dependence is free from all prior assumptions about its form.
Since the standard way of producing probability distributions (by making histograms) only works well for dense data, mutual information uses a method based on *kernel density estimation* [68]. These probability distributions are then used to form the mutual information, $I(x, y)$ [67].

Figure 5.2 contains the mutual information coefficients for the Icelandic model after initial thresholding i.e. with 30 variables. This variable selection step is combined with the PLSR step to avoid any redundant variables being included in the model that will

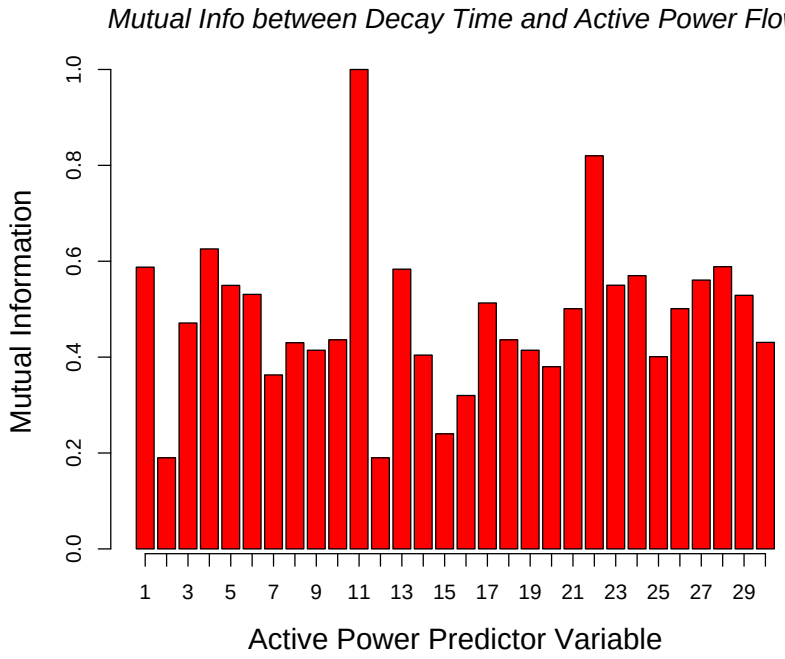


Figure 5.2: Mutual Information between mode decay time constant and MW predictors for the Icelandic 0.21Hz Mode

inhibit prediction capability. The top ranking mutual variables are compared with the principal variables from the PLSR latent vectors to eliminate relatively insignificant variables. In general, the top ranking variables in both PLSR and MI tend to agree with one another i.e. the top ranking mutual variables also occur in the PLSR latent vector.

$$I(X,Y) = \sum_{X}\sum_{y}\left[P(x,y)log_2\left\{\frac{P(x,y)}{P(x)P(y)}\right\}\right] \tag{5.8}$$

The mutual information equation is shown in equation 5.8. The mutual information is high if one data series provides a large portion of information about the other and low if it provides little. Input variables can thus be selected in a multivariate problem by deriving $I(x,y)$ for each of them in conjunction with the response variable and picking those for which this value is largest. If both signals are normally distributed, $I(x,y)$ reduces to correlation and provides identical results to a correlation analysis [69]. Figure 5.3 shows how the two variable selection methods compare to each other. It is interesting to note that the PLSR step retains all the most significant mutual information variables such as variables 11,23 and 24. Some of the higher correlated variables in the MI step such as variables 6 and 16 are removed. This process is

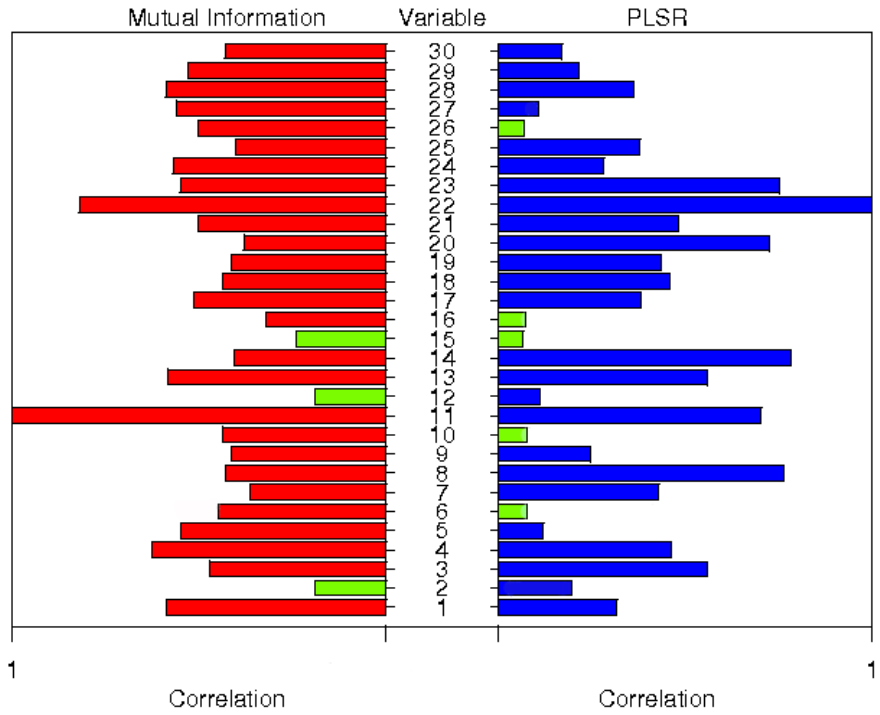


Figure 5.3: Comparison of Mutual Information and PLSR Regression. Green bars show redundant variables

illustrated in figure 5.3 where variable 10 (GB19) has a reasonable MI coefficient but is not included in the significant PLSR latent vectors. This means that variable 10 must be collinear with variable 9 (GB18) and is thus removed from the regression as for a given change in both GB18 and GB19, GB18 explains a greater portion of the response variable variance in conjunction with the latent vector, even though they have comparable regression coefficients.

It has been shown that there is generally an agreement between the mutual information and PLSR step. Although the coefficients are quite different in some places, all the important information is retained during both procedures. The PLSR step addresses the multicollinearity problem while the mutual information highlights the important non-linear correlations. As a result all the important variables are retained whilst some of the redundant variables have been removed.

5.3.3 Wavelet Variable Selection Methods

Given a regression model in terms of wavelet coefficients, it is interesting to establish which scales are the most important for developing a good regression model with the highest prediction capability. By finding such predictive-selective scales, we are focusing on certain band-limited frequency regions that are important for the prediction of the dependent variable. These important scales will therefore represent the underlying features responsible for a successful prediction.

Optimal Scale Combinations

One way to find important scales is to test the total number of all possible scale combinations for their predictive ability. The binomial in equation (5.9) describes the number of combinations that exist for selecting i different scales from a total number of K scales. Let i range from 1 to K which in turn produces K binomials. The total number of scale combinations will thus be the sum of the K binomials:

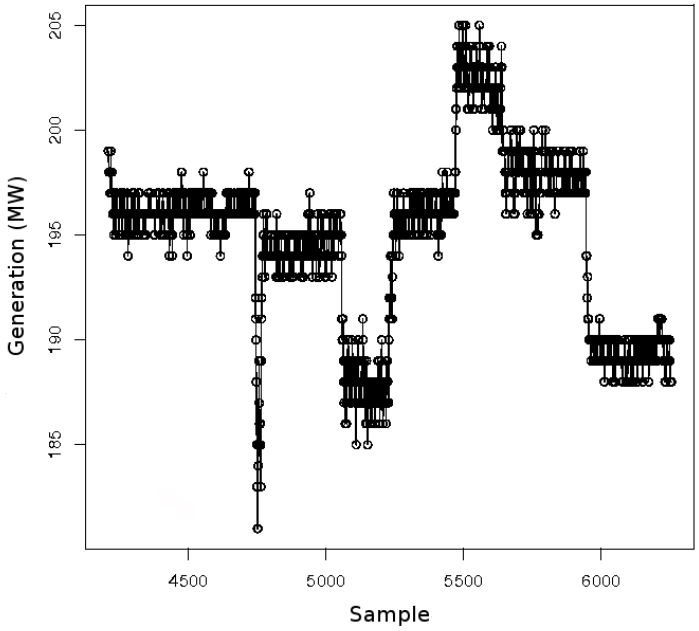
$$2^k - 1 = \sum_{i=1}^{K} \binom{K}{i} \tag{5.9}$$

In total there are 2^K-1 different scale combinations for K scales. If we had three scales [1 2 3] we could make a representation of the original data set using the following scale combinations: [1], [2], [3], [1 2], [1 3], [2 3] and [1 2 3]. Let us index each such combination as $c_1, c_2, \ldots, c_N$ where N=2^k-1. Associated with each c_j there is a RMS value r_j derived from applying the regression model on the validation set using the representation dictated by the scale combination in c_j. It is now possible to sort the duplets (c_j,r_j) with respect to r_j. Regression models with with low values of r_j are the only models of interest as the low value indicates a good model fit.

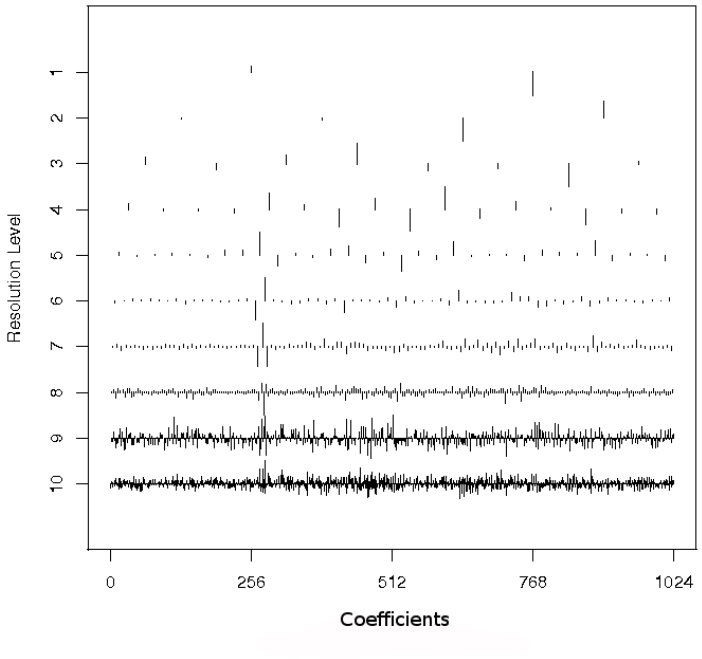
Signal Characteristics

Figure 5.4 displays a typical active power signal obtained from a well known power system compared with the wavelet coefficients of the signal derived from a discrete wavelet transform. This active power signal seems to be quite erratic with very high frequency changes throughout. Signals such as this can be "cleaned" by thresholding

the wavelet coefficients derived from the signal. Figure 5.5 again shows the comparison

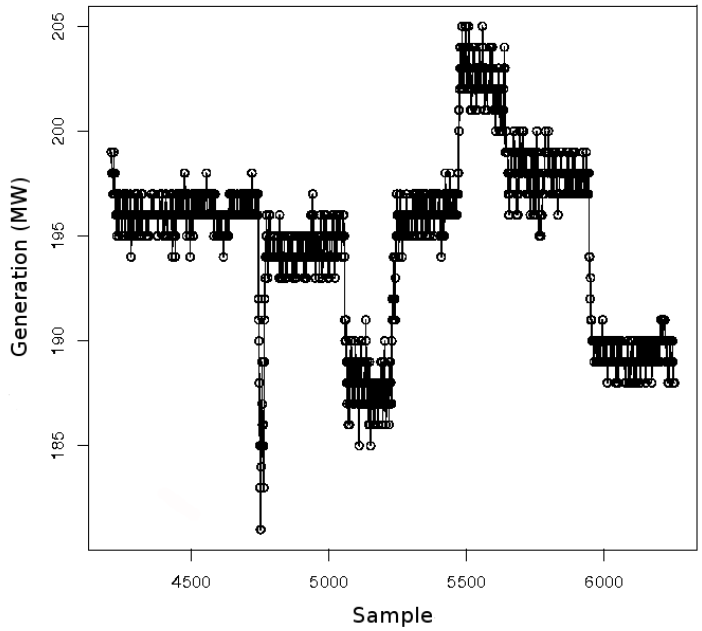


(a) Active power signal

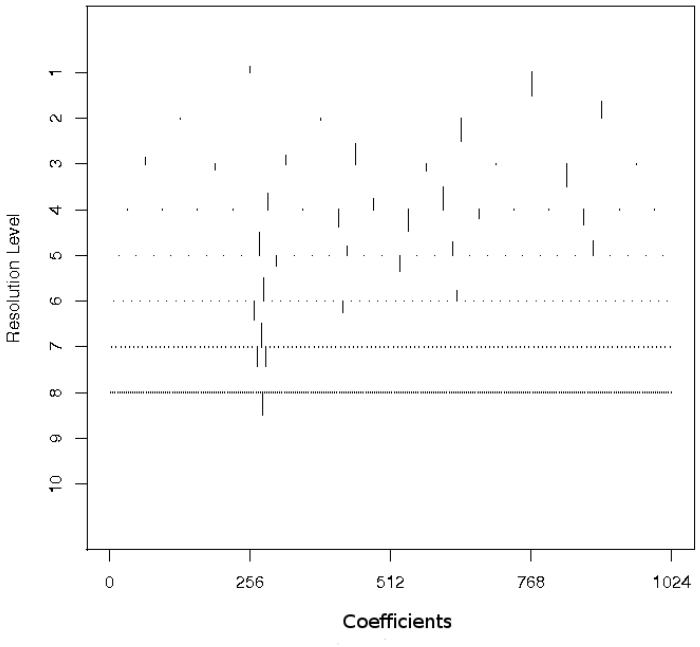


(b) Deconstruction

Figure 5.4: Comparison of active power signal and its discrete wavelet transform



(a) Active power signal



(b) Deconstruction

Figure 5.5: Comparison of active power signal and its thresholded discrete wavelet transform

between the active power signal and a thresholded version of the wavelet coefficients. The "hard" thresholding technique (see section A.3.3) has been used in this case to remove any high frequency low amplitude dynamics across the higher resolution levels. This allows the underlying (low frequency) dynamics of the signal to be retained without distortion, which can be used in event detection or subsequent regression analysis.

It can be seen how resolution levels 9 and 10 have been removed as these correspond to the high frequency bands of the signal that contain the very low amplitude changes. The remaining coefficients can then be used to reconstruct the signal without this distortion. This is effectively done by superimposing all the remaining wavelets that are left after thresholding. These wavelets are defined by their resolution level as well as their scaling and wavelet coefficients. Therefore, these translated and scaled copies that are determined by the scaling and wavelet coefficients are recombined at each resolution to reconstruct the original signal.

The scaling coefficients correspond to the smooth underlying data and the wavelet coeffients represent the sharp high frequency changes that have been superimposed. This is why the wavelet coefficients are used in the statistical analysis as they represent sharp changes in system dynamics. The reconstruction can be seen in figure 5.6 in comparison to the original active power signal.

It is possible to see how the reconstructed signal (red) tracks the original signal (black). However, the reconstructed signal does not contain any high frequency components that resemble noise, as the high frequency components have been thresholded and subsequently removed from the analysis. The clean signal remaining contains information on large signal changes as opposed to small, noisy changes in the active power. This may be useful in further analysis as the statistical models can be improved by removing certain frequency components that contain little information. In both figures, sharp changes coincide with large wavelet coefficients across most of the resolution levels. There are some changes in the signal that are represented by large wavelet coefficients across a smaller number of frequency bands and these represent less complex dynamics that are influenced by a smaller range of frequencies. Some disturbances in the signal, like the minimum value spike shown in figures 5.4 and 5.5 are constructed from a wide

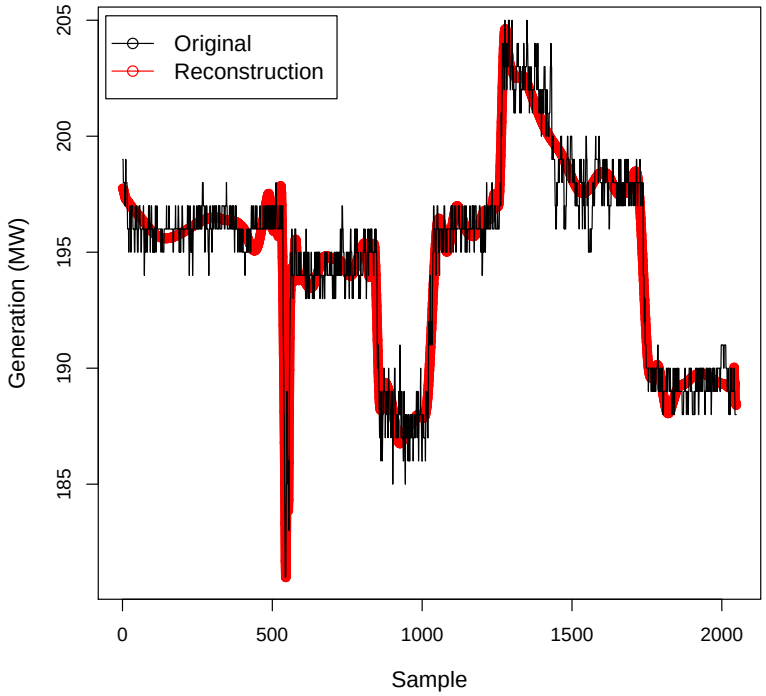


Figure 5.6: Reconstruction of Power Signal using only the hard thresholded coefficients

range of frequencies and are therefore more complex in their makeup as well as their effect on the response variable.

These examples of deconstructed signals allow us to see the benefit of using wavelet transforms in analyzing dynamic performance. By determining the frequency at which most of the system dynamics occur we can then limit our analysis of system variables and mode dynamics to the discrete frequency bands where most of the interesting dynamics take place. This essentially allows the removal of any redundant frequencies that have little effect on the mode being analyzed. This streamlining increases the prediction capability of the statistical model under consideration. This could be used to provide more accurate information on the dynamics of the system and provide a more informative guide on the causes of poor mode damping.

5.4 Generalized Linear Models

5.4.1 Overview

In statistics, the generalized linear model is a flexible generalization of ordinary least squares regression. The generalized linear model generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value. The link function provides the relationship between the linear predictor and the mean of the distribution function. In the case of the binomial distribution the link function is given as follows:

$$X\beta = ln\left(\frac{\mu}{1-\mu}\right) \tag{5.10}$$

This is known as the *logit* function, the plot of which is shown figure 6.1.

In a generalized linear model, each outcome of the dependent variable, Y, is assumed to be generated from a particular *distribution function* from the exponential family, that includes a range of probability distributions such as normal, binomial and Poisson distributions, among others. The mean, μ, of the distribution depends on the independent variables, X, through:

$$E(Y) = \mu = g^{-1}(X\beta) \tag{5.11}$$

where $E(Y)$ is the expected value of Y; $X\beta$ is the linear predictor, a linear combination of unknown parameters, β; g is the link function. In this form, the variance is typically a function, V, of the mean:

$$Var(Y) = V(\mu) = V(g^{-1}(X\beta)) \tag{5.12}$$

It is thus convenient if V follows on from the exponential family distribution, but it may simply be that the variance is a function of the predicted value. The unknown parameters, β, are typically estimated with maximum likelihood, maximum quasi-likelihood or Bayesian techniques [70].

The generalized linear model consists of three elements detailed as follows:

- A distribution function f from the exponential family
- A linear predictor $\eta = X\beta$
- A link function g such that $E(Y) = \mu = g^{-1}(\eta)$

These three attributes will be discussed in the following subsections.

Distribution Function

The exponential family of distributions are those probability distributions, parameterized by θ and τ, whose density functions f (or probability mass function for the case of a discrete distribution) can be expressed in the form:

$$f_Y(y;\theta,\tau) = exp\left(\frac{a(y)b(\theta) - c(\theta)}{h(\tau)} + d(y,\tau)\right) \tag{5.13}$$

where τ, the dispersion parameter, is typically known and is usually related to the variance of the distribution. The functions a, b, c, d and h are known. Many, although not all common distributions are in this family.

The variable θ is related to the mean of the distribution. If a is the identity function, then the distribution is said to be in canonical form. If, in addition, b is the identity and τ is known then θ is called the canonical parameter and is related to the mean through:

$$\mu = E(Y) = c^{'}(\theta) \tag{5.14}$$

Under these conditions it can be shown that [71]:

$$Var(Y) = c^{''}(\theta)h(\tau) \tag{5.15}$$

Linear Predictor

The linear predictor is a quantity which incorporates the information about the variables into the model. The Greek symbol η is typically used to denote a linear predictor. It is related to the expected value of the response variable through the link function.

The variable η is expressed as linear combinations of unknown parameters β. The coefficients of the linear combination are represented as a matrix of independent variables X. Thus, η can be expressed as:

$$\eta = X\beta \tag{5.16}$$

The elements of X are either measured by the experiments or stipulated by them in the modeling design process.

Link Function

The link function provides the relationship between the linear predictor and the mean of the distribution function. There are many commonly used link functions and their choice can be somewhat arbitrary. It can be convenient to match the domain of the link function to the range of the distribution function mean.

When using a distribution function with a canonical parameter θ, a link function exists which allows for X^TY to be a sufficient statistic for β. This occurs when the link function equates θ and the linear predictor. Table 5.2 contains all the canonical link functions and their inverse's (mean function μ) which are used for several distributions in the exponential family. In the case of the exponential and gamma distributions,

Distribution	Name	Link Function	Mean Function
Normal	Identity	$X\beta = \mu$	$\mu = X\beta$
Exponential	Inverse	$X\beta = \mu^{-1}$	$\mu = (X\beta)^{-1}$
Gamma	Inverse	$X\beta = \mu^{-1}$	$\mu = (X\beta)^{-1}$
Inv. Gaussian	Inv. Sqrd	$X\beta = \mu^{-2}$	$\mu = (X\beta)^{-1/2}$
Poisson	Log	$X\beta = \ln(\mu)$	$\mu = \exp(X\beta)$
Binomial	Logit	$X\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{1}{1+exp(-X\beta)}$

Table 5.2: Canonical link functions in generalized linear model

the domain of the canonical link function is not the same as the permitted range of the mean. In particular, the linear predictor may be negative which would give an impossible negative mean. When maximizing the likelihood, precautions must be taken to avoid this. An alternative is to use a non canonical link function.

5.4.2 Fitting a generalized linear model

Maximum Likelihood

In section 5.3.2 we discussed the use of mutual information to determine how well non-parametric variables are correlated with one another. In fact, both mutual information and the maximum likelihood are connected as demonstrated by equation (5.17) which denotes the expectation of the log-likelihood:

$$\frac{1}{T}E\{L\} = \sum_{i=1}^{n} E\{log(f_i(w_i^T x))\} \tag{5.17}$$

If f_i were equal to the actual distributions of $w_i^T x$, the first term of the equation would then be equal to $-\sum_i H(w_i^T x)$. Thus the likelihood would be equal, up to an additive constant, to the negative of the mutual information as given in equation (5.8). In practice, the connection is even stronger as the distributions of the independent components are not known. A reasonable approach in this case would be to estimate the density of $w_i^T x$ as part of a maximum likelihood estimation method and use this as an approximation of the density of the s_i i.e. a potential predictor variable. The likelihood and mutual information are, for all practical purposes, equivalent. However, there is a small difference between them as the estimation of the densities f_i must be correct. If the information on the nature of the independent components is wrong the maximum likelihood estimation will give spurious results [72].

However, maximum likelihood is a very popular method used for fitting statistical models to data and providing estimates for model parameters. For a fixed set of data and underlying probability model, maximum likelihood picks the values of the model parameters that makes the response data more likely than with any other values of parameters. Maximum likelihood estimation gives a unique and easy way to determine a solution in the case of a normally distributed dataset as well as many other distributions.

In the case of generalized linear model's, the maximum likelihood estimates can be found using an iteratively reweighed least squares algorithm: either a Newton-Raphson

method with updates of the form:

$$\beta^{(t+1)} = \beta^{(t)} + \mathcal{I}^{-1}(\beta^{(t)})u(\beta^{(t)}) \tag{5.18}$$

where $\mathcal{I}(\beta^{(t)})$ is the observed information matrix (the negative of the Hessian matrix) and $u(\beta^{(t)})$ is the score function, or a Fisher's scoring method:

$$\beta^{(t+1)} = \beta^{(t)} + \mathcal{T}^{-1}(\beta^{(t)})u(\beta^{(t)}) \tag{5.19}$$

where $\mathcal{T}(\beta^{(t)})$ is the Fisher information matrix. If the canonical function is used, then the two methods are the same. In terms of Bayesian fitting techniques it must be noted that the posterior distribution cannot be found in closed form and must be approximated using Laplace approximations or some type of Markov chain Monte Carlo method [71].

5.4.3 Examples of generalized linear models

General Linear Models

The general linear model (GLM), which is a broad statistical model in itself, may be viewed as a case of the generalized linear model with an identity link function. Remember that the link function provides a relationship between the linear predictor and the mean of the distribution function. Therefore, the identity link function provides the relationship ($X\beta$) between the predictor and a normally distributed dependent variable [73]. This model will be used later in this chapter with wavelet transformed data with the model being scored in terms of deviance as opposed to RMSE.

Statistical Deviance

In statistics, deviance is a quality of fit statistic for a model that is often used for statistical hypothesis testing. The deviance for a model is defined as:

$$D(y) = -2[log(p(y|\hat{\theta}_0)) - log(p(y|\hat{\theta}_s))] \tag{5.20}$$

Here, $\hat{\theta}_0$ denotes the fitted values of the parameters in the model M_0, while $\hat{\theta}_s$ denotes the fitted parameters for the "full model": both sets of fitted values are implicitly functions of the observations y. Here the full model is a model with a parameter for every observation so that the data is fit exactly. This expression is simply -2 times the log-likelihood ratio of the reduced model compared to the full model.

Suppose in the framework of the GLM, we have two nested models, M_1 and M_2. In particular, suppose that M1 contains the parameters in M2, and k additional parameters. Then, under the null hypothesis that M_2 is the true model, the difference between the deviances for the two models follows an approximate chi-squared distribution with k-degrees of freedom. Note that here both models M_1 and M_2 would be subsets of the full model used to define the zero of the deviance criterion [71].
The degrees of freedom dictate the number of variables in the model that are free to vary. In the case of the null model, all the variables are free to vary but in the reduced model the degrees of freedom are dependent on the number of parameters in that model. In other words, the null model has $n-1$ degrees of freedom where n is equal to the number of components in the model or the data length. The degrees of freedom of the reduced model will have $n-1-p$ degrees of freedom where p is the number of β parameters. Thus the null model and reduced model are both submodels of the full model with a parameter for every observation and zero degrees of freedom.

In summary, the general linear models, which are essentially linear regressions on normally distributed wavelet transformed data, are fitted using maximum likelihood and scored using statistical deviance. These GLM's are used later in this chapter where a variety of models are tested and scored for comparison, in order to determine the most accurate results.

Linear Regression

A very important example of a generalized linear model, which is also an example of a general linear model is linear regression. Here the distribution function is the normal distribution with constant variance and an identity link function which is the canonical

link if the variance is known. Unlike most other generalized linear models there is a closed form solution of the maximum likelihood parameter estimates.

Binomial Regression

When the response variable, Y, is binary i.e. taking on values of 1 or 0, the distribution function is generally chosen to be the binomial distribution. The interpretation of the distribution function mean, μ_i, is then the probability p, of Y_i taking on the value of 1. There are several link functions available for binomial functions with the most typical being the canonical *logit* link:

$$g(p) = ln\left(\frac{p}{1-p}\right) \tag{5.21}$$

A generalized linear model with this setup is known as a logistic regression model. In addition, the inverse of any continuous cumulative distribution function (CDF) can be used for the link function since the CDF's range is [0,1] i.e. the range of a binomial mean. The normal CDF Φ is a popular choice and yields the *probit* model with a link function given as:

$$g(p) = \Phi^{-1}(p) \tag{5.22}$$

The identity link is also sometimes used for the binomial data to yield the linear probability model. However, a drawback of this model is that the predicted probabilities can be greater than one or less than zero. It is possible to fix these non-sensical probabilities that lie outside of [0,1] but interpreting the coefficients can be difficult. The models primary merit is that near $p = 0.5$, it is approximately a linear transformation of the probit and logit.

The variance function of the binomial data is given by:

$$Var(Y_i) = \tau\mu_i(1-\mu_i) \tag{5.23}$$

where the dispersion parameter τ is typically fixed at exactly one. When it is not one, the resulting $quasi-likelihood$ model is often described as binomial with overdispersion or $quasibinomial$.

Count Data

Another example of generalized linear models includes *Poisson regression* which models count data using the *Poisson distribution*. The link is typically the logarithm which is the canonical link. The variance function, given below, is proportional to the mean:

$$Var(Y_i) = \tau\mu_i \tag{5.24}$$

where the dispersion parameter, τ, is again typically one. When it is not, the resulting $quasi - likelihood$ model is often described as Poisson with overdispersion or *quasipoisson* [74].

Quasilikelihood Models

As mentioned previously, if the data being used to develop a model is overdispersed then it can be difficult to determine what distribution, if any, should be used in the generalized linear model. In this case a *quasi* model can be used which doesn't assign any particular distribution to the data but allows the variance to be expressed as a function of the response variable mean. In this way, the non-constant variance can be modelled and a quasi-likelihood estimator based on generalized least squares can be used to estimate the model coefficients.

5.4.4 Extensions of the generalized linear model

Generalized Additive Models

Generalized additive models (GAMs) are an extension of generalized linear models in which the linear predictor η is not restricted to be linear in the covariates X but is the sum of smoothing functions applied to the predictors so the GAMs equation looks like:

$$\eta = \beta_0 + f_1(X_1) + f_2(x_2) + \ldots + f_m(x_m) \tag{5.25}$$

The smoothing functions $f_i x_i$ may be fitted using parametric or non-parametric means and are estimated from the data, thus providing the potential for better fitted models than other methods [75].

5.4.5 Statistical Weighting

Statistical weighting is useful for estimating the values of model parameters when the response values have differing degrees of variability over the combinations of the predictor values. In parameter estimation, as in regular least squares, the unknown values of the parameters $\beta_0, \beta_1, \ldots, \beta_n$ in the regression function are estimated by finding the numerical values for the parameter estimates that minimize the sum of the squared deviations between the observed responses and the functional portion of the model. Unlike least squares, however, each term in the weighted least squares criterion includes an additional weight, w_i, that determines how much each observation in the data set influences the final parameter estimates. These weights can be used in GLM models too as they have the ability to modify predictor contributions to the response variable which may enhance the prediction capability [76].

5.5 Developing a Statistical Model

It has been proposed that the sensitivities of oscillatory modes to static parameters can yield information on operational variables that most strongly influence particular dynamic conditions. This information can be used in the following ways:

- To tune the analytical dynamic model to replicate the measured behaviour
- To determine suitable control actions that can be taken in real-time when the near instability is detected
- To identify one or more plants that significantly degrade the dynamic performance of the system

Since power system oscillations are essentially interactions between interconnected dynamic systems, there can be several different factors that combine to cause a poorly damped or unstable condition. As discussed in earlier chapters, generator active and reactive power output, loading conditions, interarea transfer levels and network topology can all play a part in creating poorly damped network conditions.

A power system operator will have access to SCADA data that defines the operating

condition of the system in terms of active and reactive power flows in the network. If the dynamic model of a power system was perfect, the system state could be estimated from the SCADA measurements and the dynamic measurements derived from the eigen-analysis of the system. However, since models are never perfect there can be poorly damped conditions that are not reflected in the model. In this case, a conventional model-based analysis is of limited use. Showing that mode damping can be related to active or reactive power flow in a particular generator feeder or interconnector helps power system operators and analysts determine an appropriate course of action to alleviate any instability.

This section presents some preliminary experiments investigating the correspondence between the statistical model and the analytical dynamic model. A wide range of random system loading scenarios were generated using a number of system models. The use of statistical methods to identify the factors that most strongly affect the behaviour of various modes will be investigated. The conclusions from the statistical modeling of the system can be assessed using conventional small-disturbance analysis techniques. The feasibility of describing the dynamic state of a model network using a statistical process is shown. The same statistical process can be applied to real system measurements and can provide valuable insights into the factors that influence troublesome oscillatory dynamics.

In these models, active and reactive power generation and flow are used as they supply the oscillation energy to the system and give an indication of the reactive power support that can sustain oscillations. Close to the stability limits, damping is non-linearly related to active and reactive power and the relationship becomes more linear further away from the stability limits. Since active and reactive power both share a non-linear relationship with voltage, the relationship between damping and voltage is further distorted and as a result, it is not used in this study.

In addition, for small systems active and reactive power variables are sufficient as oscillations can be explained in terms of a small number of generators even if they are

interarea in nature. For larger systems, bus voltages and angles may incorporate information that may be of use for more complex modes that are caused by a larger number of generators across the system.

5.5.1 Statistical Modelling of the Power System

The procedure for generating a statistical model that describes the dynamic behaviour of the system in based on a number of different statistical analyzes were mentioned previously. Essentially, all the processes derive some sort of multivariate statistical relationship between the active and reactive power flow variables and mode damping. Having derived these relationships, statistical measures can be used to determine how closely these statistical models match the observations. Furthermore, the statistical models can be used to identify the factors that most strongly influence the damping of a mode.

Damping is measured in either mode decay time constant or in damping ratio. It was noted in a previous study [2] that the relationship between active and reactive power flow and mode damping ratio was approximately linear for most of the operational range but becomes more non-linear at the limits of the operational range i.e. close to the stability limits. It was also noted in [2], that in cases where damping is close to the stability limit (or crosses it), damping ratio is a better response variable. However, a study of response variable types is given later in this chapter in terms of GLM's and wavelets. It was found that decay time constant is the best response variable to use with the wavelet transform. This is due to increased correlation between the larger wavelet coefficients, as decay time constant boasts a greater variance than damping ratio.

5.5.2 Procedure for Statistical Analysis

The following procedure was used to convert and tabulate the data for use in subsequent statistical steps. The steps below are given for a simple linear regression although the data will be subject to the same procedure for any other step in the source location methodology.

- Select a mode from the system and express damping in both Mode decay time constant (τ) and Damping Ratio (ζ).
- Select a suitable timeseries during which an event occurs and make sure there is sufficient shift in the data to account for delay in dynamics output (real-time data only).
- Build a matrix of independent (predictor) variables comprising data series of active and reactive power as well as voltage and line flows.
- Filter data and select valid observations. Negative or undamped cases may be included if ζ is used as a damping expression.
- Perform variable selection procedures on the data to determine useful predictors.
- When using wavelets make sure the data length is a power of two to allow suitable reconstruction and deconstruction.
- In some cases, but not all, it may be useful to normalize the predictor matrix and this will be explicitly stated when done.
- Normalizing the data negates the effect of larger variables on the response variable so that all the independent datasets can be compared on the same scale. This may not be desirable as larger generators tend to effect damping more that smaller ones.
- Apply the wavelet transform to both the decay time constant, damping ratio and system state data.
- Perform linear regression or GLM analysis on both mode decay time constant or Damping Ratio depending on which performs better.
- Determine the variables that significantly influence the prediction by calculating deviance statistic for the the GLM.
- General linear models are fitted using maximum likelihood and scored using deviance.

This procedure is standard across every statistical procedure performed whether it is a GLM or a logic regression step. The Gaussian GLM or general linear model which is used later is exactly the same as linear regression except the model is fitted using maximum likelihood and scored in terms of deviance as opposed to RMSE. The above procedure will be extended in Chapter 6 to cover logic regression. In the case of mutual information, mutual coefficients are produced by comparison of each predictor matrix column to the response vector containing the damping values. This is similar to how PLSR is performed on the vectors (response) and matrices (predictors) to produce combinations of variables (factors) that best explain variation in the Y-space.

The Y-Space is defined as the n-dimensional space containing n dependent variables with the m dimensional X-Space containing m independent variables. A PLSR model tries to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. When the data has been processed into the aforementioned matrix format, variable selection procedures and thresholding etc. may be performed to simplify the analysis.

In the case of linear regression and general linear models, variables with a relatively large regression coefficient that were deemed to have a substantial effect on mode damping were retained and highlighted. In these cases a positive test of significance (p-test) and t-test (used to determine differences in means) were also performed to determine if the results were significant and that their means i.e. response and predictor means, were sufficiently different relative to the spread of their variability. If variables with relatively large regression coefficients pass both the p and t-tests then they are considered important factors that influence the mode damping.

In the case of GLM's the parameter estimates are divided by their standard error to give a z-value. However, these results are derived from asymptotic theory where the distribution is limited as the sample size increases. Due to this unreliability, a suitable model is confirmed by a deviance test which determines if a reduced model is significantly different from the full model (one that has a parameter for every observation).

A model with a relatively small deviance statistic, in comparison to the full model, is one that has has a good fit with the data which allows important predictors to be extracted and used in further analysis.

The strongest predictor variables may not correspond directly to the true casual factors that directly influence dynamic behaviour. However, it is likely that the true casual factors are strongly correlated with the predictors. Using linear regression it is not possible to distinguish between a true casual factor and a variable that is strongly correlated to it but variables that are well correlated with the good predictors and are significant (p-test) may be included in further analysis. If these variables are multicollinear and have little effect on the response then they will be removed from the model via PLSR [2].

The development of GLM's and the subsequent Logic Regression methodology allows the true casual factors to be determined from the model. As we will see, the casual combinatorial factors are not always well correlated with the strong predictor variables or highlighted in the GLM's, due to the interactions between predictors and their effect on the response variable. With the use of mutual information and PLSR, the initial set of predictors can be reduced with the removal of insignificant predictors. These steps should improve the model prediction capability and further highlight static parameters that influence certain dynamic conditions.

5.6 Data Preprocessing and Thresholding

Due to the nature of real-time power system data it is necessary to preprocess the data to filter and threshold it before it is used in any statistical models. Electromechanical modes are constantly changing in the system with frequency, amplitude and damping values shifting in response to changes in system conditions. When a mode that is being tracked suddenly deviates in frequency or becomes unobservable, the decay time constant is no longer measurable and a NaN (not a number) is reported. Due to this, many decay time constant signals reported from real-time dynamics software are

discontinuous. System state measurements suffer a similar problem in that NaN's are reported when a signal cannot be measured. These anomalies in system measurements cannot be used in the source location methodology as much of the statistics used are sensitive to missing values and may give spurious results if included.

In the case of the system state variables, NaN's were simply replaced with 0's as NaN's were predominantly reported for deactivated units in the real systems. Any NaN that was reported from an active unit was interpolated to produce a continuous flow of data as it is unlikely that a unit is deactivated for a few sample periods of the system state estimator. The decay time constant variables were also interpolated as a smoother dataset produced better results. The next section contains details on the interpolation process as well as other pre-processing steps.

5.6.1 Data Extraction

The approach to the source location problem in this thesis uses the extraction of dominant mode parameters from a real system using small-disturbance oscillations measured by synchronized phasor measurements at various points in the system. The phasor measurement unit (PMU) outputs are processed with specialized software to continuously derive stability indicators such as mode frequency, decay time constant, damping ratio, amplitude and mode phase. These measurements derive from small perturbations always present in the system. There is a delay between the dynamic outputs and the initial receipt of the signal in question as the algorithm takes time to accumulate data to calculate the dynamic values. This delay means a time-shift has to be included in the statistical models so that the dynamic values and the system state variables are correctly aligned. This is only required with real data as model data is already aligned due to the instantaneous nature of eigenvalue analysis. When enough of the dynamic signal has been collected, the algorithm outputs dynamic information at a rate of 0.33Hz or 0.05Hz i.e. 1 output every 3 seconds or 20 seconds.

An energy management system (EMS) that includes supervisory control and data acquisition (SCADA) systems are used to acquire information on the system state. These

measurements obtained via transducers in the power system are updated every five seconds and can be used to determine the causes of troublesome dynamics via a statistical model. This data is normally sampled at a rate that allows it to be compared to the dynamic data. For example, the data may be sampled to produce a value every minute. The 20 second dynamic information would also be downsampled to produce an output every minute i.e. downsampled by 3. The dynamic data would then be shifted back (due to the delay) to align the datasets. A shift of half the delay time is used as their is some overestimation in the algorithm. Once the data has been pre-processed into this format it can then be used in subsequent variable selection steps and statistical models.

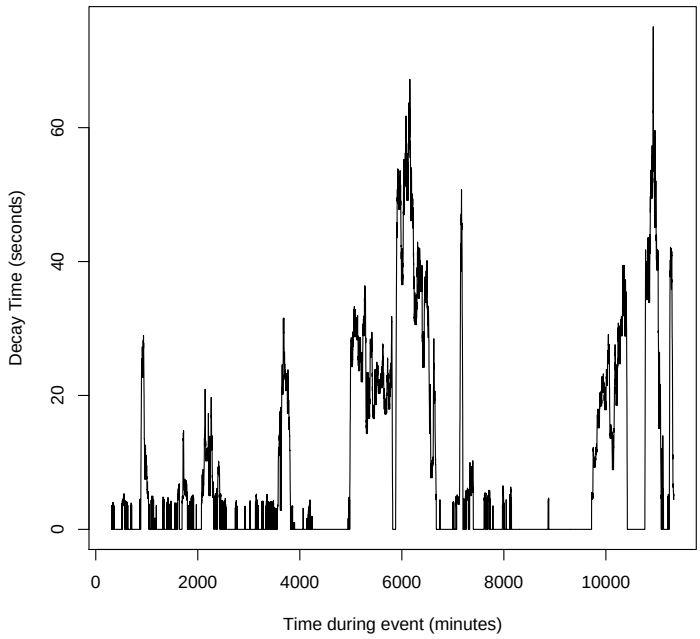
5.6.2 Interpolation

Figure 5.7 shows a comparison of an interpolated decay time constant record in comparison to the raw decay time constant record. The interpolation allows the data to be used in the GLM's and the logic regression steps without the consequence of spurious results being produced due to a large population of NaN's in the data. It may also serve to enhance the statistical models and allow significant effects to be highlighted as the presence of NaN's and zeros may give false negative correlations that are not present in the system.

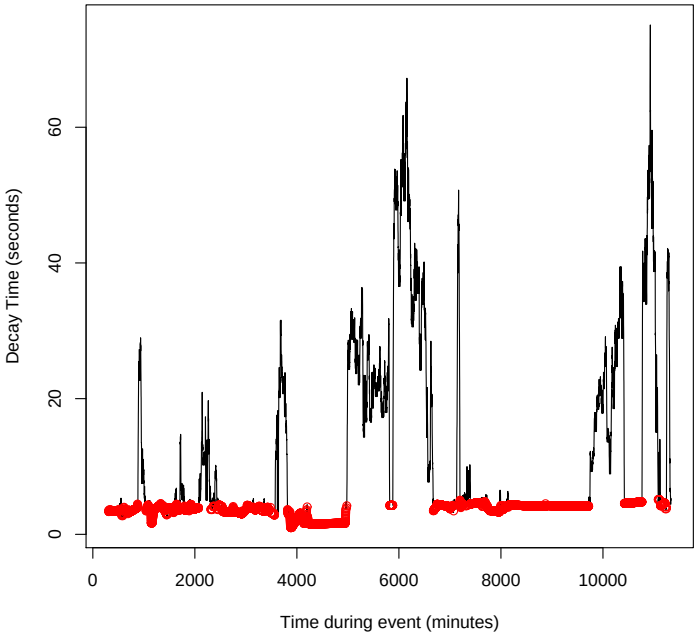
5.6.3 Variability and Thresholding

Since a typical power system is subject to constant perturbations, the measured system variables such as active and reactive power usually have a high variance. Coincident changes in both system state and damping are of particular interest in this study as comparison of these coincident changes can indicate the important variables that are strongly linked to mode damping.

Due to the length of events and the variability of power system data, periods of constant values in the interesting signals are not salient as the signals are quite dynamic. Even though a constant generator MW output may have an effect on mode damping it is not likely to stay constant during the sampled window lengths used in this study.



(a) Raw, non-interpolated 0.48Hz decay time constant signal from Australia



(b) Interpolated 0.48Hz decay time constant signal from Australia

Figure 5.7: decay time constant profile with (a) interpolation and (b) without interpolation

As a result, all the data from the real system and system models were thresholded, both in terms of absolute mean value and the variance of the data. This step filtered out any low value signals with low variance that, by their inactivity, would be very unlikely to have an effect on damping. The variance and thresholding limits were not constant and were changed manually depending on the characteristics of the system generation. Figure 5.8 shows an example of a MW signal from the Icelandic model.

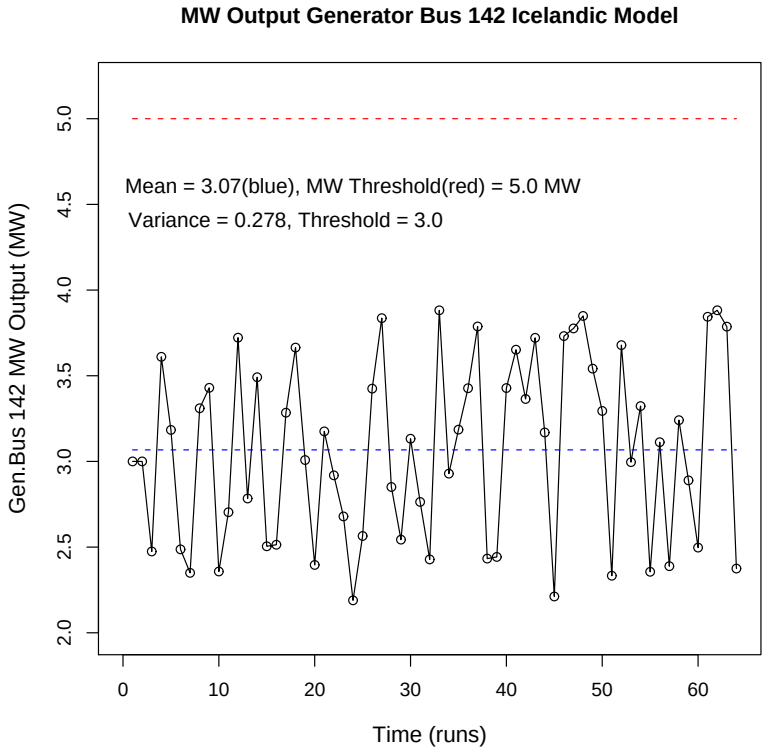


Figure 5.8: Generator Bus MW Output with mean and variance thresholding illustrated

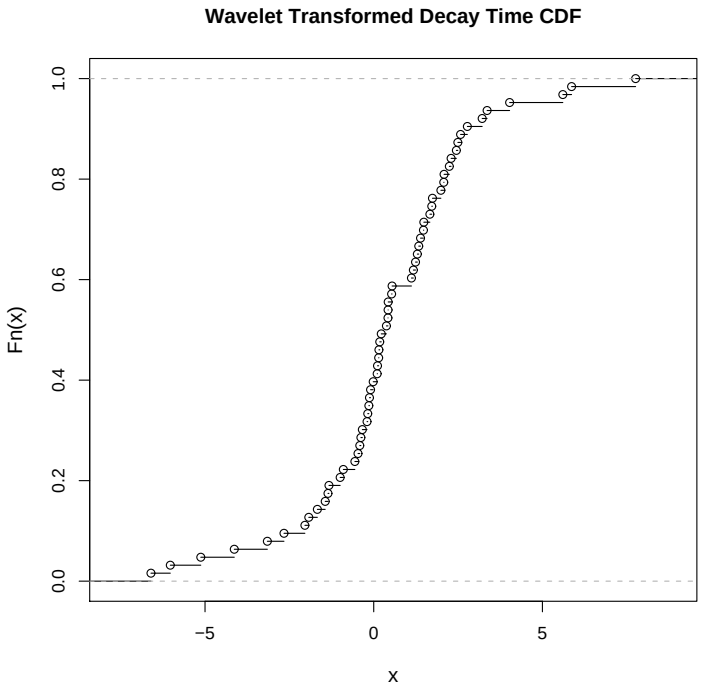
The red dashed line represents the absolute mean threshold while the blue line is the mean value of the signal. The variance of the signal is 0.278 while the threshold is set at 3.0. The variance corresponds to the mean of the square of the deviances from the expected value. A value of 3.0 would be the required variance in the signal to cause significant changes in the damping profile. Thus, both the mean and the variance of the signal have to be below a certain threshold to be omitted. This ensures that low mean signals with large power swings and low variance high mean signals are retained as these types of signals can have an effect on mode damping.

5.7 Results from the General Linear Model

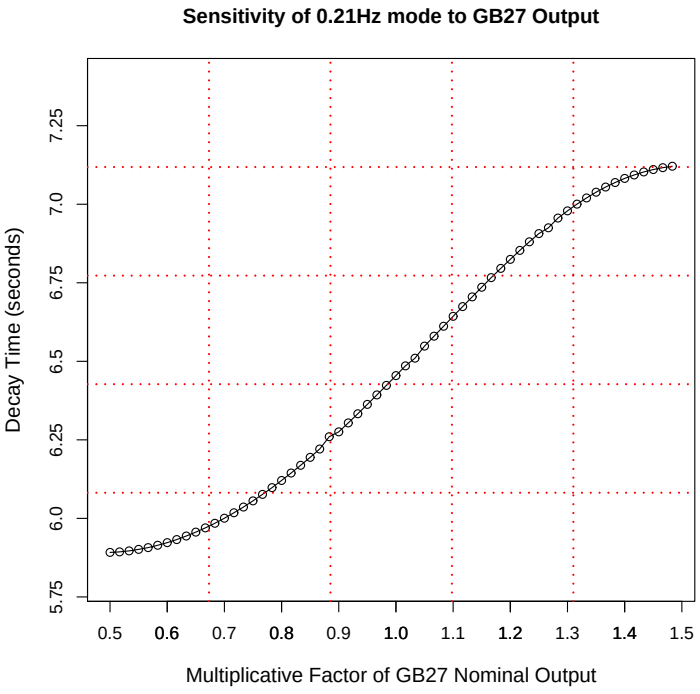
Figure 5.9(a) is a plot of the cumulative distribution function of the wavelet transformed 0.21Hz Icelandic mode decay time constant. It can be seen that the data follows what looks like a normal distribution or what can best be described as quasi-normal. The CDF could also suggest a gamma or Poisson distribution if count data with the correct parameters were used to define the distribution. In section 5.4.1 we discussed the use of distribution families within generalized linear models that could be used to model the distributions of desired response variables. In the case of raw mode decay time constant and damping ratio the distribution is not known, therefore it would be unwise to commit to a distribution within the exponential family. In this case, a $quasi-distribution$ family within the generalized linear models or a Gaussian GLM (general linear model) can be used with the wavelet transformed data as it has the effect of normally distributing the data. The general linear model is essentially a linear regression scored in terms of deviance.

The *quasi* family, as its name suggests, allows the ambiguous distributions to be modelled along with a link function. In section 5.4.1 we discussed how a link function could be used to describe the relationship between the predictor and the mean of the response variable distribution. Figure 5.9(b) shows how the response (decay time constant) relates to an increase in a predictor which in this case is generator G8 in Iceland (see figure 4.2). The resulting function looks like a logistic sigmoid function. In this case a logit link function may be used but the CDF of the raw damping values (both ζ and τ) is highly erratic and discontinuous, which is turn gives unacceptable results when they are used in a general linear model.

However, the wavelet transformed response variables have a more linear dependence on the wavelet transformed predictor variables. Even though the CDF in this case is slightly erratic it is suitable for the GLM family and looks much more like a real normal distribution than the raw case. Figure 5.10 shows the relationship between the wavelet transformed values of decay time constant and generator G8 MW output



(a) Cumulative Distribution Function of Wavelet Transformed decay time constant data



(b) decay time constant record with GB27 ramped up from 0.5 to 1.5 Nominal Output

Figure 5.9: Response CDF and sensitivity to predictor variables

during normal system operation without any ramping of the generator G8 output i.e. it is not a sensitivity analysis. The relationship is more linear due to the fact that large changes in raw values produce a disproportionate change in the corresponding wavelet coefficients i.e. the relative change in wavelet coefficients is much less and as a result the relationship approaches linearity. It can be seen that generator G8 (which is

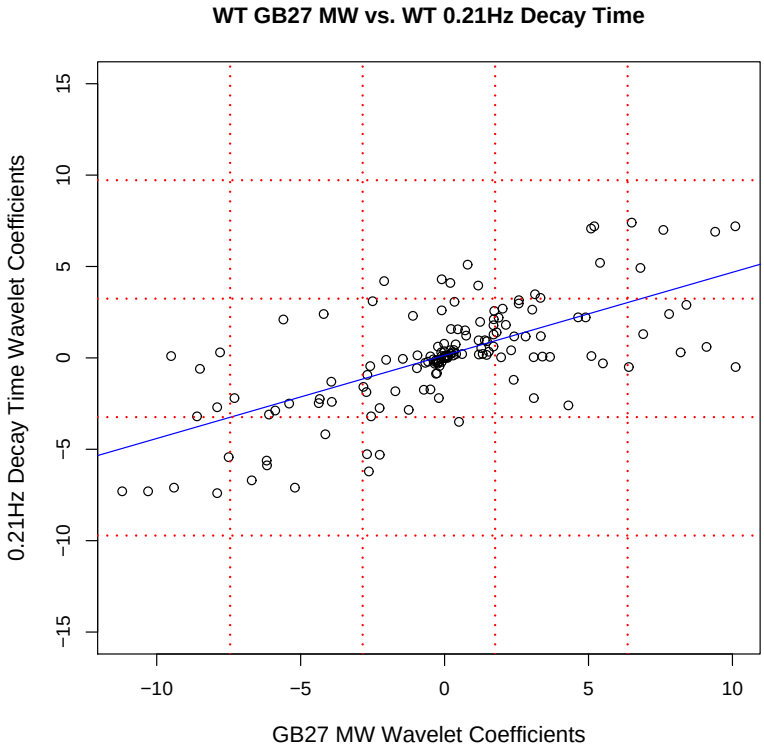


Figure 5.10: Wavelet Transformed decay time constant versus Wavelet Transformed GB27 MW Output.

connected to generator bus GB27) is closely linked to the mode decay time constant as under normal operating conditions, a large change in G27 wavelet coefficient produces an increase in decay time constant wavelet coefficient. This suggests that when there is an increase in G27, decay time constant seldom decreases and when it does, the decrease is minimal.

In this case, the identity link function can be used to model the wavelet transformed relationship as GB27 MW output is well correlated with the mode decay time constants. This means that both signal changes are aligned and will thus have corresponding large

wavelet coefficients (for the most part) which improves the linearity of the relationship. It is due to this wavelet-related linearity that a general linear model can be used with the data to give good model prediction.

5.7.1 Development of GLM Expression in the 16 Machine Case

Initially the wavelet transformed general linear model (WGLM) was tested using the 16 machine model that was introduced in chapter 4. The model was set up with the random permutations and intentional adustment to induce some poorly damped events. In this case window sizes of 64 and 128 were used as these were of sufficient length to allow enough events to occur so that a reliable statistical model could be developed. The wavelet transformed model, as well as the various other statistical models were developed in the statistical programming language R. A maximum data length of 256 has been imposed in this study as this should be long enough to produce well fitted models while avoiding unnecessarily long computation times. This is especially true with power systems because they tend to evolve over time and different combinations of predictors begin to affect damping. By spreading a window over a long time, the regression coefficients can become smudged and it can be difficult to interpret the results. This serves to inhibit the prediction capability of the model and its ability to identify strongly correlated predictors.

In the 16 machine case study, the active power generation, decay time constants and damping ratios were used to determine significant predictors that can be used to control mode damping. The wavelet transform was used on both the response and predictor variables as decay time constants and damping ratios are system properties that change and these changes are determined by system states such as active and reactive generation. A study of the relationship between raw decay time constant/damping ratio and wavelet transformed system predictors was also carried out to determine the best optimal scale combination and response variable to use i.e. decay time constant or damping ratio.

5.7.2 Model Setup

As mentioned in chapter 4, eigenvalue analysis was performed for each operating point determined by the state variable adjustment. For each loadflow scenario the eigenvalues, frequency, decay time constant and damping ratio of the system modes were tabulated along with other system states, including active and reactive power generation. These variables were selected as they can be used to explain corresponding dynamic behaviour of particular modes. The eigenvalues for the 16 Machine 0.43Hz interarea mode are shown in figure 5.11, with frequency on the y-axis and the attenuation constant on the x-axis.

The aim of this section is to develop a linear model from wavelet transformed variables to yield a model that explains the dynamic behaviour of the system. It is also known that the signals contain redundancy i.e. some frequency bands contain little or no interesting dynamics. By using optimal scales of wavelet frequency bands it is hoped that the model can be further improved to enhance predictability as well as preserving and highlighting significant predictors that can be used to influence damping.

5.7.3 Interarea Mode Tracking

Of particular interest in figure 5.11 is the spread of the eigenvalues in response to changes in system conditions. This figure only displays the extreme eigenvalues of the nominal 0.43Hz mode, which shows the range of frequencies and decay time constants across the loadflow scenarios. The frequencies of the nominal 0.43Hz mode lie between 0.38Hz and 0.53Hz which gives a spread of 0.15Hz. In order to use any source location methodology based on a system model, the modes in question have to be validated. In other words, it has to be shown that the interarea modes under study can be faithfully tracked when the system conditions are perturbed. By doing this, we ensure that the correct mode attributes are used in the methodology as opposed to different modes of similar frequency. The methods used to track modes have been detailed in [50] but will be summarized here for convenience.

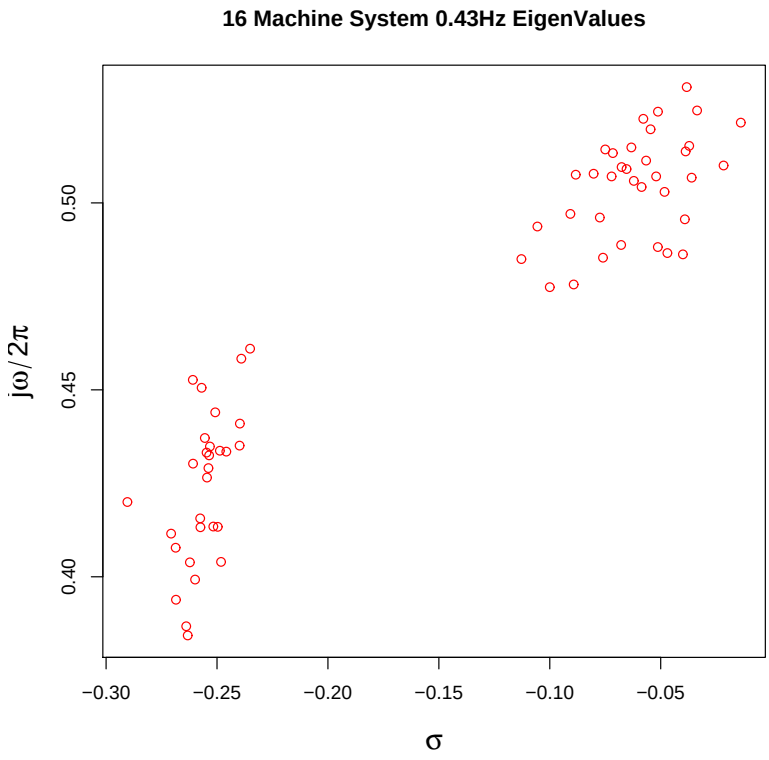


Figure 5.11: EigenValue plot of 0.43Hz interarea mode showing the extreme eigenvalues in terms real and imaginary components

Just like the real system, any complex power system model generates a large number of modes as different loadflow scenarios are run. In order to validate a model it is of vital importance that the correct mode from the system and model are compared. It is also vital that the correct mode is tracked over time in both the model and the real system [50].

Due to the large number of modes and the spread of the eigenvalues in the model, it is necessary to cluster the modes to determine the movement of a particular mode under different system loading conditions. The eigenvalues can be clustered across the loadflow cases according to:

- Mode Frequency
- Generator Participation
- Mode Observability

Thus, a mode can be identified as being consistent between different system conditions if the mode frequency is similar, the mode is most observable in the same locations and if the same set of generators participate in the recurring modes.

As we have seen in figure 5.11 the frequency spread across the eigenvalues was around 0.15Hz. In relative terms this is quite a large shift from the nominal value of 0.43Hz so subsequent modes must be validated in terms of participation factors and/or observability. Figure 5.12 displays subplots of generator participation factors for the same mode that occurs at different frequencies across the frequency range. This is for the 16 Machine 128 data length case whose single line diagram is shown in figure 4.5. It is clear that the mode is consistent in terms of participation factors across the frequency range. In all cases, generators 14 and 15 which are aggregated generators rated at 10000MVA, can be seen to be of high participation, while a number of less significant machines such as generators 6,7 and 16 are also reported. These participation patterns across the 0.15Hz range show that the mode has successfully been tracked and is suitable for further analysis.

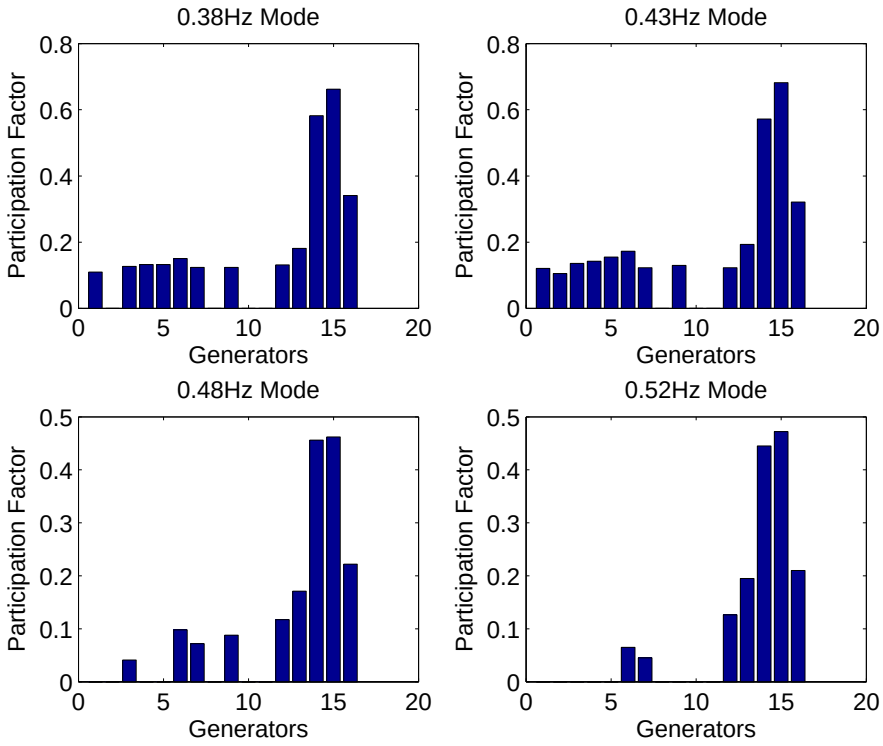


Figure 5.12: Comparison of Participation of Generators in the 0.43Hz (nominal) Mode

Generator 13 is the slack generator which absorbs or generates any power imbalance in the system. It can be seen that G13 participation factor is relatively small which means that the imbalance in the system isn't large. In come cases, when the slack generator is at high output, certain modes look like they are dependent on slack generation. However, this is misleading as the mode(s) will actually be dependent on load and generation patterns elsewhere in the grid that manifest themselves through the slack generation. By having a well balanced model, more realistic case studies can be generated and as a result the slack generation usually has a relatively low participation factor. In addition, by plotting the mode plots for the nominal 0.43Hz mode and the extreme 0.52Hz mode, it can be deduced that the nominal 0.43Hz mode has been tracked to 0.52Hz and that both modes exhibit similar phases and amplitudes for the most observable machines (especially G14 and G15). There is some discrepancy in mode phase and amplitude which can be attributed to a relatively extreme change in system conditions. By comparing both the mode plots and participation factors for

the 0.52Hz mode we can be confident that the mode has been successfully tracked.

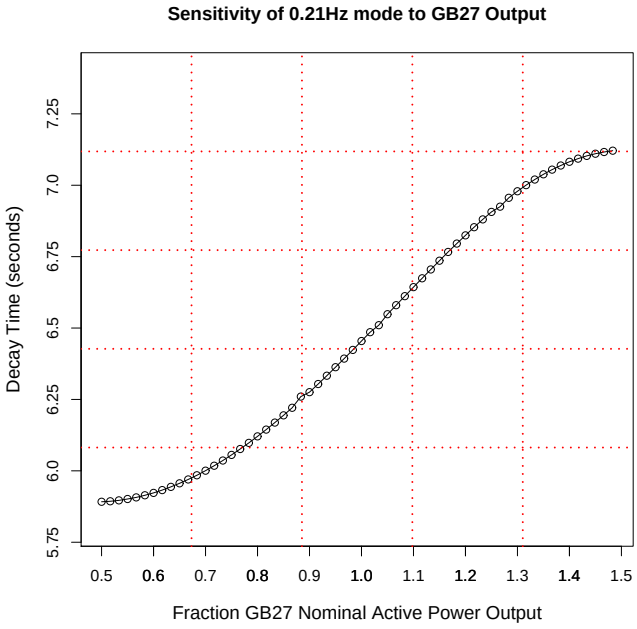
5.7.4 decay time constant versus Damping Ratio

In a number of studies [51], both damping ratio and decay time constant are used as the damping predictor. In this subsection a study is presented in which both decay time constant and damping ratio are used to determine the best statistical model to provide the best predictions. Figure 5.13 contains a comparison of the decay time constant and damping ratio sensitivities to GB27 MW output for the 0.21Hz Icelandic mode. It can be seen that both cases have a non-linear relationship with MW output especially near the generator limits. Both the graphs exhibit a logistic and inverse logistic relationship between GB27 MW output and decay time constant/damping ratio respectively. However, the damping ratio seems to have a slightly less logistic (slightly more linear) relationship with the predictor than the decay time constant case. This may be more useful in a raw linear regression analysis but it is still a non-linear function so the resulting fit may be poor.

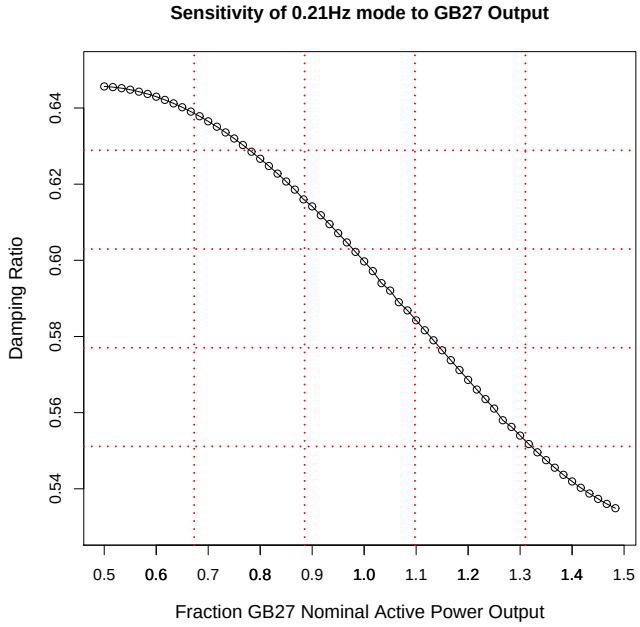
The relative overall change in the damping ratio is smaller than the relative change in decay time constant. In terms of the wavelet transform, the larger changes in the decay time constant signal in turn produce larger wavelet coefficients while maintaining linearity. These larger relative changes in the transformed response variable works well in the general linear model to produce well fitted models. The increased variation of the linear response of the wavelet transformed decay time constant allows it to be used more successfully in general linear models and nullifies any advantage of the damping ratio. In addition, a link function can be specified to model any non-linear relationship between the predictors and the mean if this is required.

Table 5.3 shows the percentage of variance explained (calculated from deviances) for the 16 Machine and Icelandic models of various data lengths with the same variable selection procedures applied to both decay time constant and damping ratio models.

It can be seen how decay time constant performs slightly better in the 16 Machine 128



(a) Sensitivity 0.21Hz Icelandic Mode to GB27 MW Output (decay time constant)



(b) Sensitivity 0.21Hz Icelandic Mode to GB27 MW Output (Damping Ratio)

Figure 5.13: Comparison of Damping versus GB27 MW Output in terms of decay time constant and Damping Ratio

Model	Data Length	τ Variance (%)	ζ Variance (%)
16MC	64	63	63
16MC	128	73	72
Ice	128	61	62
Ice	256	71	70

Table 5.3: Comparison of WGLM deviances (shown in percentage variance explained) with using both damping ratio and decay time constant as response variables

and Iceland 256 length models while the damping ratio performs better in the Iceland 128 length model. The larger variation in the decay time constant values (defined as 1/real part of the eigenvalue) results in a marginally better prediction from the wavelet transformed models over the range system conditions and models tested. These results combined with the fact that the decay time constant is independent of mode frequency (see eqns. 2.20 and 2.21) has led to the use of decay time constant as the response variable in the rest of this thesis.

Frequency independence is important as this study is only interested in the stability of low frequency modes. Damping ratio is approximately inversely proportional to mode frequency, so low frequency modes will have higher damping ratio magnitudes. By using decay time constant, a direct frequency independent stability indicator can be obtained. The use of decay time constant gives very similar results to the damping ratio and this has been observed over all the models tested. As a result, little or no information is lost in choosing decay time constant over damping ratio. However, the damping ratio could also be used as it is virtually identical to decay time constant in terms of performance.

5.7.5 Development of GLM expression

Figure 5.14 shows a plot of the actual (simulated) decay time constant values for the 16 machine case versus the predicted values derived from a general linear model consisting of the raw decay time constant and active power. The actual values are derived from the simulated system via eigenvalue analysis and the predicted values derive from the statistical models developed using the simulated decay time constant and loadflow variables.

The actual model refers to The blue line in figure 5.14 depicts the perfect model fit with a gradient of 1.0. The best fit line can be seen to be quite different from the perfect line in figure 5.14 and the variance of the prediction is very large and thus very inaccurate. The variation in the response variable explained by the model only amounts to 28% which means that over two-thirds of the model variation is unaccounted for in the model and thus cannot be attributed to any variable or groups of variables in the model. This explained variance (exp. var.) is shown in figure 5.14 and will also be included in all subsequent prediction plots as a means of scoring the model using statistical deviance.

In figure 5.15, the predicted wavelet transformed (WT) decay time constant is plotted

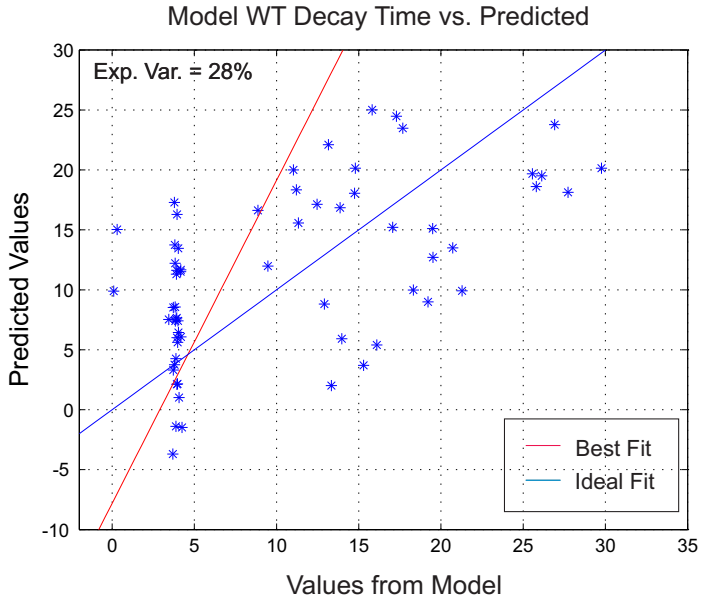


Figure 5.14: Predicted (statistical model) Raw Data versus Actual (simulated) Raw Data from a GLM with 16 Machine 64 Length decay time constant

against the actual WT decay time constants for the 0.43Hz interarea mode. A best fit line, as well as a perfect fit line has been included to show any linear relationship that may be present. The predicted values are derived from the general linear model. In this case, the full wavelet spectrum is used from the dataset of length 64 which corresponds to 63 wavelet coefficients. There is one coefficient less as the wavelet transform

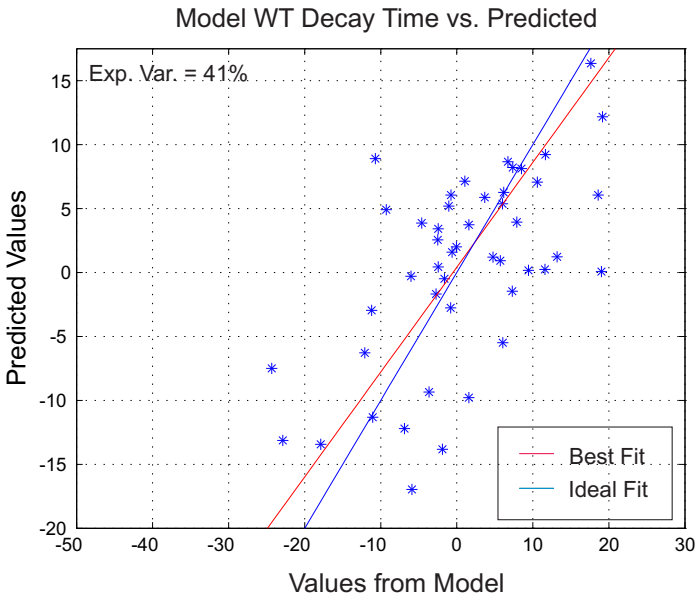


Figure 5.15: Predicted Wavelet Coefficients versus Actual Wavelet Coefficients from 16 Machine model for data length of 64

is based on the pyramidal algorithm which decomposes the signal into binomial powers i.e. 1,2,4,8....2^n. Therefore the total number of variables is the sum of these levels which is always one less that the data length.

The prediction in this case is not acceptable as there is too much residual error in the model. The null deviance of the general linear model is 2756.4 on 63 degrees of freedom with the residual deviance measuring 1617.3 on 48 degrees of freedom. These deviances suggest that the current full wavelet spectra model only explains, (2756.4-1617.3)/2756.4= 0.41, 41% of the variability of the response variable. The linear fit is quite close to the blue ideal fit but clearly this isn't an acceptable result. This is due to the high variance of the statistical prediction relative to the ideal prediction which indicates poor prediction as seen in figure 5.15. The estimated regression coefficients will be erroneous and will not be well correlated with the response variable. As a result, important predictors that control mode decay time constant cannot be highlighted with confidence. However, by comparing figures 5.14 and 5.15 the benefits of using wavelets can be clearly seen in the increased linearity of the model.

The above values used to calculate the deviance scores are derived from maximum likelihood estimators (MLE) which score the fit of the dependent variable in terms of probability. For a given set of predictor parameter values the MLE attempts to maximize the likelihood or probability that the dependent variable will achieve the actual response value. The MLE function is maximized i.e. the set of predictor parameters that produce the highest MLE score are chosen as the model parameters. In the above case, the MLE score for the full and reduced model were 2756.4 and 1617.3 respectively which resulted in 41% explained variance. In other statistical models the MLE magnitude output may vary but the deviance is scored in relative terms between the full and reduced model.

The degrees of freedom are the number of values in the final calculation of a statistic that are free to vary. In the case of the full wavelet model, residual degree of freedom is 48 which corresponds to 63-15 i.e. the number of observations minus the number of parameters. Fifteen variables were used in the model as the slack generator was omitted.

Figure 5.16 shows a plot of wavelet coefficients derived from the discrete wavelet transforms of actual and predicted decay time constants of the 0.43Hz mode. In this case only the wavelet coefficients from the two highest resolution levels are used. These resolutions correspond to signal dynamics at frequencies of $32f_b$ and $64f_b$ where f_b is equal $1/data\ length$. Therefore, for a 64 length window $32f_b = 0.5$Hz frequency scale, which corresponds to 32 wavelet coefficients describing changes in the signal that happen between two samples. Some of the 15 variables that were removed (redundant frequency resolutions) were outliers that caused the statistical model to be distorted. By using selected frequency scales a better fitted model may be attained as the outliers no longer affect the prediction.

With the optimal frequency scales used in the general linear model the null deviance is 1994.5 on 48 degrees of freedom with the residual deviance measuring 738.6 with 33

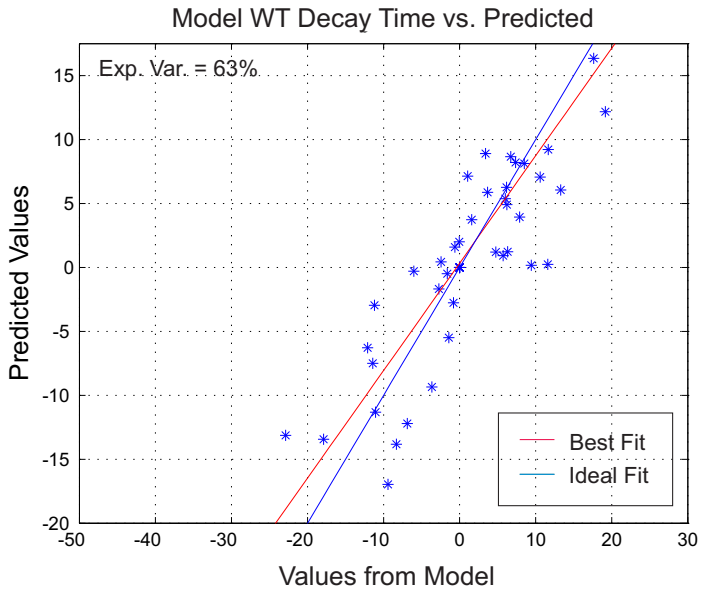


Figure 5.16: Predicted Wavelet Coefficients versus Actual Wavelet Coefficients from power system model with two highest resolutions

degrees of freedom. This corresponds to 63% of the response variability being explained by the model. Again, this is an unacceptable result but it nevertheless highlights the benefit of using optimal wavelet scales with the linear model.

Figure 5.17 shows the prediction accuracy when 128 datapoints are used in the GLM.

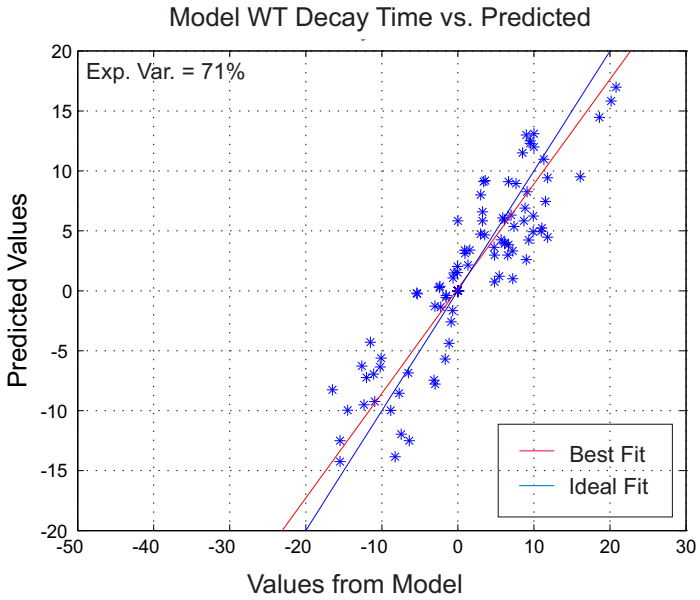


Figure 5.17: Predicted Wavelet Coefficients versus Actual Wavelet Coefficients from model for 128 length dataset with three highest resolutions

For this model, frequency resolution levels of 4, 5 and 6 were used which correspond to frequency component values of $32f_b$, $64f_b$ and $128f_b$. From here on in f_b will be measured in Hz for convenience sake and we will assume that a loadflow is solved and eigenvalue analysis is performed every second. The null deviance of the model is 3324.3 on 111 degrees of freedom with the residual deviance measuring 963.4 on 96 degrees of freedom. These results mean that 71% of the response variable variance is explained by the GLM. This result is much better and can be used in a source location study as the predictors can be tested for residual error and significance.
It can be seen that by using the discrete wavelet transform in conjunction with the general linear model, good results can be produced with acceptable residual error. In the following sections, variable selection procedures outlined previously will be used to see if even better results can be obtained.

5.8 Sensitivity Analysis

5.8.1 Sixteen Machine Model

This section presents a selection of results from a sensitivity analysis that was conducted to determine how the 0.43Hz system mode reacted to changes in those significant predictors obtained from the statistical analysis. The significant variables are shown in table 5.4. The sensitivity analysis is akin to comparing participation factors as the behaviour of the eigenvalues in response to changes in system state determine the participation factors. However, participation factors are not readily available in real systems due to the low penetration of PMU's. In a sensitivity analysis, by changing a system state i.e. active power at a generator, it is possible to view the dynamic effects directly as opposed to relying on participation factors.

Participation factors are explained in section 2.7 with the defining equation given in equation in 2.23. Participation factors are a measure of the relative participation of of the kth state variable in the ith mode and vice-versa and can be used to determine suitable controls for electromechanical modes as well as determining suitable locations for PMU placement. Just like a linear regression model, the β parameters were calcu-

Variable	Parameter Value(β)	Residual Error	Norm PF
Gen 6	0.0663	0.015	0.355
Gen 7	0.0615	0.011	0.321
Gen 9	-0.0288	6.7e-03	0.189
Gen 14	0.0534	9.7e-03	0.822
Gen 15	0.0234	4.7e-03	1.00
Gen 16	0.0227	5.4e-03	0.525

Table 5.4: Parameter values, Residual Error and Normalized Participation Factors for Selected Variables from 128 point optimal scale GLM

lated for the independent wavelet transformed variables that were used as independent predictors. The most significant variables from the 128 length optimal scale GLM are shown in Table 5.4 with participation factors averaged over all the loadflow runs.
It must be stressed, that the β coefficients represent the gradients of the relationships between the decay time constant wavelet coefficients and the system state wavelet coefficients, not the raw values as in conventional regression analysis.

The results show that certain system variables contribute more than others when predicting the response. They suggest that to adequately predict the damping of the 0.21Hz mode, the generation at generators 6,7,9,14,15 and 16 are required as these have the highest β parameters. Since these variables are required for a good prediction they must exert some influence on mode damping. This means that these active generators can be used to control damping during a poorly damped event or may be used to avoid any initial instability.

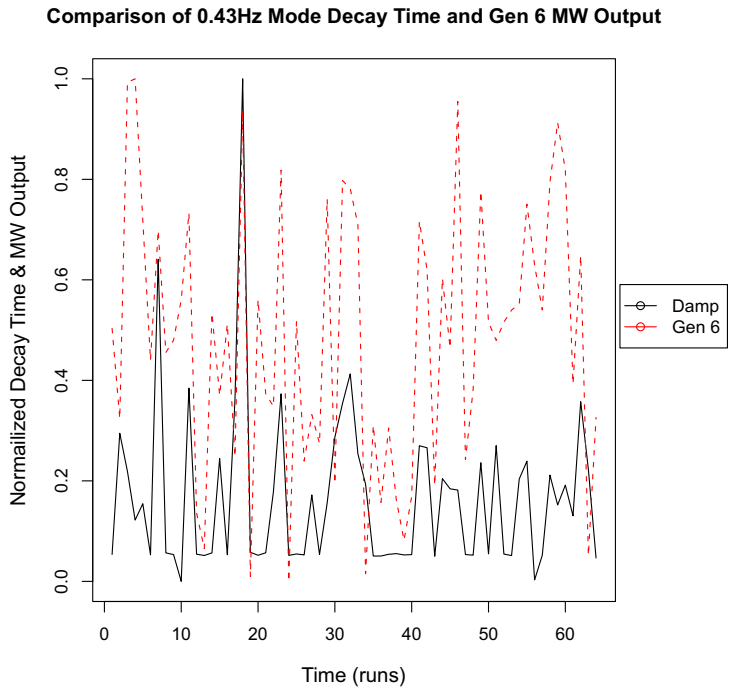
The fourth column of table 5.4 contains the normalized average participation factors derived from the 16 Machine model over each loadflow scenario. In figure 5.12 the participation factors across the mode frequency range were plotted and it was noted that generator 14,15,16,6 and 7 were the highest participating machines. Generator 13 is the slack generator and can be seen in figure 5.12 to have a relatively low participation factor. However, generator 13 was omitted from the GLM and will be ignored from here on when referencing the participation factors. As a result, the participation factors in table 5.4 have been normalized to the maximum value which is attributed to

generator 15.

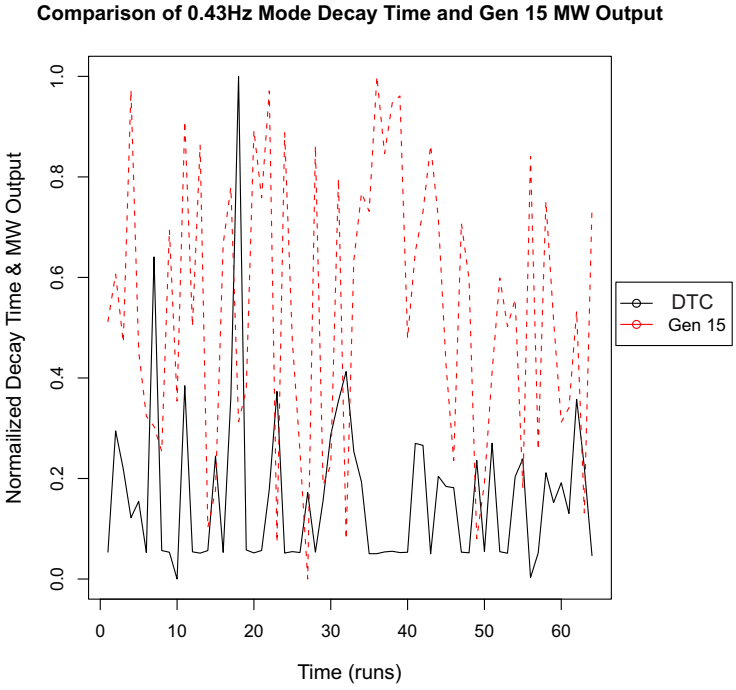
Column 4 shows that there is some disagreement between the participation factors and the significant statistical variables. From the wavelet transformed general linear model, generators 6 and 7 were reported to have the highest regression coefficients. They also have appreciable participation factors though not the maximum values which are attributed to G14, G15 and G16. G14 and G15 have reasonable regression coefficients but again they are not the most significant variables derived from the wavelet transformed GLM (WGLM). Figure 5.18 shows the comparison of the most significant variable from the WGLM versus the most significant variable derived from the average participation factors for the 0.43Hz mode. The comparisons in the figure detail the decay time constant over half the 128 datapoint window, as plotting the whole window would render the comparisons indistinguishable. Generator 15 is the highest participating machine and its MW output has been plotted along with generator 6 MW output in figure 5.18.

It can be seen that G6 is a better predictor than G15 as the majority of the decay time constant peaks coincide with peaks in G6 output. The same cannot be said for G15 as it seems to have an ambiguous relationship with mode damping. At certain points they seem to coincide whereas at other instances they are out of synchronism. In comparison, the participation of G6 in the 0.43Hz mode was 0.355. This example shows that the wavelet transformed GLM's have the ability to extract statistical relationships not detected in an algebraic study. A similar result was found in [2].

Although the model is not perfect it can still be used in conjunction with participation factors to determine the best predictors to use for damping control. In this case, the participation factors do not produce a result as satisfactory as the WGLM as evidenced in the superpositions of Generators 6 and 15 with decay time constant. In other cases, the participation factors do produce predictors that are well correlated with decay time constant and as such, can be used to control damping. This is the case with G14 (see table 5.4 and figure 5.12) as this is a truly significant predictor.
However, a statistical method is also required to highlight strong predictor variables



(a) Comparison of 0.43Hz Mode decay time constant and Gen 6 MW Output



(b) Comparison of 0.43Hz Mode decay time constant and Gen 15 MW Output

Figure 5.18: Comparison of Gen 6 and Gen 15 MW Output with 0.43Hz Mode decay time constant for section of 128 length window

not highlighted in a participation factor analysis.

5.8.2 16 Machine Wavelet Transformed GLM with Variable Selection

In section 5.3.2 we discussed two methods of variable selection that allowed the removal of insignificant or redundant variables from the statistical models. In this section both mutual information and PLSR will be applied to the 16 machine data to see if any redundant variables can be excluded. The same variable selection procedure will then be applied to the Icelandic model to see if a better fitted model can be attained. Figure 5.19 shows the coefficients derived form a mutual information and PLSR model

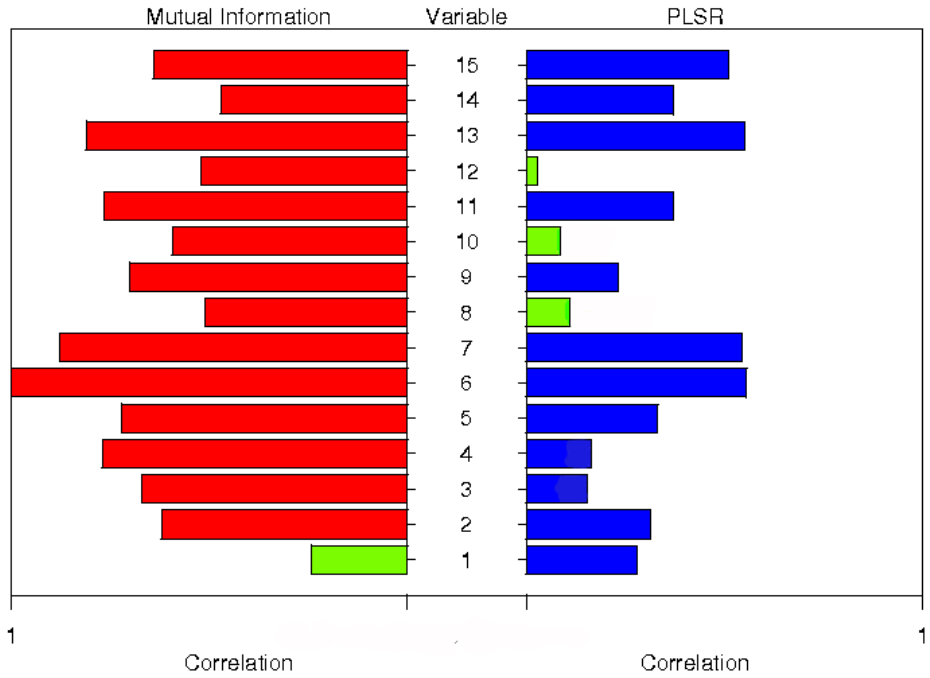


Figure 5.19: Pyramid plot showing coefficients derived from mutual information and PLSR. Green bars indicate redundant variables.

for the 16 machine data with 128 data points. The green bars indicate redundant variables that are excluded from the model. Generator 1 has been excluded as it is deemed to have negligible mutual information with the decay time constant profile. The PLSR redundant variables may have appreciable MI coefficients but may just be closely correlated with other variables.

This collinearity may give a variable the illusion of significance whereas it isn't actually

significant in terms of the PLSR latent vectors. Generators 8, 10 and 12 have been excluded as the PLSR model has deemed them to have a negligible effect in conjunction with other predictors in the model. This means that their inclusion has little effect on the response variable, however they seem significant as they are correlated to true predictors and thus have a reasonable MI coefficient.

Although both the mutual information and PLSR don't always reconcile, it is useful to combine them to remove some redundancy from the model. However, care has to be taken to ensure too many variables aren't removed as this could reduce the model prediction accuracy. Figure 5.20 shows a plot of model wavelet coeffieicents

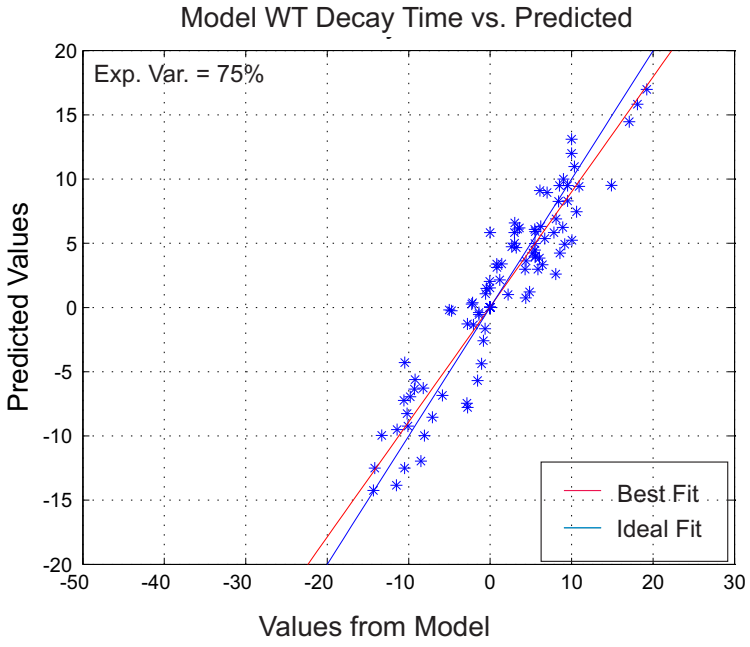


Figure 5.20: Actual wavelet coefficients versus predicted wavelet coefficients from model for variable selection model - x=y blue line shown for perfect prediction

versus predicted coefficients for the 16 machine model with 128 data points. In this case the three highest resolution levels are used again and variable selection has been implemented with generators at generator buses 1,8,10 and 12 being removed from the model. It can be seen that the model is now a slightly better fit than the same case without variable selection (see figure 5.17).

The null deviance of the model is 3113.3 on 111 degrees of freedom with the residual deviance measuring 779.9 on 96 degrees of freedom. These results mean that 75% of

the decay time constant variability is explained by the GLM. This result is slightly better than the non-variable selection case as the redundant variables serve to reduce the GLM fit by reducing the contribution of the true predictors.
These results show that with the use of variable selection the overall model can be improved further. Although the improvement in the 16 Machine case is small, the improvement in larger models with yet more predictors may be significant. This will be investigated in the next section with the use of the more complex Icelandic model.

5.9 Development of Icelandic Model GLM expression

If the strongly linked predictors from the GLM can be calculated, then a transformed linear expression can be generated which may be used to predict mode damping based on measurable power system quantities. This expression may be refined in real time as more data is made available to provide a system operator with updated information on mode dynamics. By inspection, some of the original 189 bus injection variables were removed from the Icelandic model and only generator buses were retained. This resulted in 34 variables being used in the statistical model i.e. the slack bus was removed from the original 35.
As the models used were relatively small, the active power variables were the most important as they have the greatest effect on damping due to the fact that active power output supplies oscillation energy to the system. However, reactive power will invariably contain information on the damping and also must be used in the development of the model.

In the first instance, a Gaussian general linear model was derived from all the generator active power variables which are themselves, nonlinear functions of voltage magnitudes and phase angles. Therefore, in terms of voltage magnitudes and phase angles, the derived model in nonlinear. Line flows are also functions of active generation so by using the generator output values, significant generators can be modulated to control damping. It must be noted that although the 34 variables are all active power variables, some redundancy may still exist that serves to reduce model accuracy. As a

result, some variable selection procedures were applied to the predictors later in this section to determine if the model fit and prediction accuracy could be improved.

The Icelandic 0.21Hz interarea mode was of particular interest due to the long decay time constants observed. Eigenvalue analysis was performed for each loadflow that was run using random and intentional variable adjustment. For each loadflow solution, the eigenvalues, decay time constant, damping ratio and frequency of the 0.21Hz interarea mode were recorded. In addition, active and reactive power outputs were tabulated. Power flows were also calculated but seeing as they are functions of generation, their inclusion in the model may only serve to add more dimensionality to the problem as the statistical information already exists in the generation data.
The dimensionality problem derives from the fact that multiple regression analysis is a seductive technique. We may plug in as many predictor variables as we can think of and usually at least a few of them will come out significant. A large number of variables may cause a statistical model to become less reliable as the estimates of the regression parameters may be diminished. This can lead to a lack of replication of the model if the study was repeated a number of times.

Of particular interest in the study was the correlation between the active power and mode decay time constants as the active power is responsible for supplying oscillation energy into the system. Some of the transmission line power flows in the weak part of the Icelandic network are important in an analytical study of the system as heavy loading on these lines very often leads to oscillatory instability. However, these line flows are functions of generation at Blanda and Krafla etc. (see figure 5.21) and as such these generation variables are used as the predictors as they are easiest to control.

Figure 5.22 displays the eigenvalues of the 0.21Hz mode for the simulated case with data length of 128. The eigenvalues have been plotted with frequency (Hz) on the y-axis and the attenuation constant (σ) on the x-axis. This plot shows an appreciable degree of scatter in the eigenvalues when system states are randomly changed within preset limits. Data windows of 128 and 256 were chosen to allow the wavelet transform

Figure 5.21: Top level schematic of the Icelandic Power System with PMU's

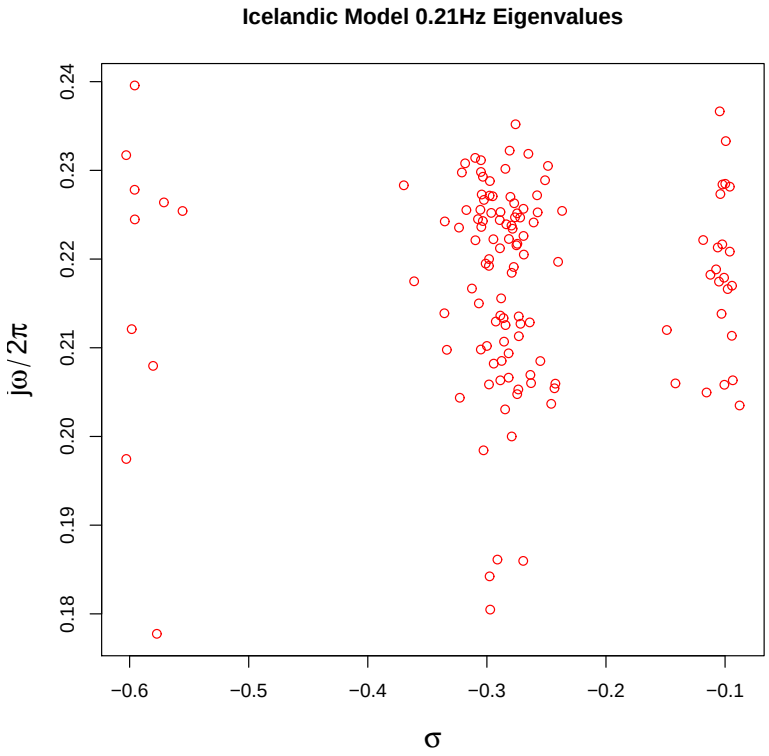


Figure 5.22: Eigenvalues of 0.21Hz interarea mode from the Icelandic Model with 128 data points

to be applied to the data, as the data length must be a power of two i.e. 2,4,64,128 to be suitable for the wavelet transform. The data length should also long enough to allow enough poorly damped events to occur so that a well fitted GLM can be developed. Once the signals were deconstructed using the DWT, a general linear model was built and the results are presented later in this section.

The β parameters were calculated for the independent wavelet transformed model

Variable	Definition	Parameter Value	Residual Error	PF
G8	Gen Bus 27	0.0834	0.0152	1.0000
G16	Gen Bus 99	0.0468	0.0123	0.8776
G17	Gen Bus 102	0.0331	7e-03	0.2113
G14	Gen Bus 52	0.0059	4.9e-04	0.0398
G19	Gen Bus 117	0.0111	2.4e-03	0.2889

Table 5.5: Parameters and Residual Error for Selected Variables from 256 point optimal scale GLM. Generator numbers shown in brackets

predictors. The most significant variables from the 256 length optimal scale GLM are shown in table 5.5 along with some of the less important predictors included for comparison. The null deviance (deviance of full model with a parameter for every observation) for this GLM is 4525.4 on 255 degrees of freedom while the residual deviance (best fit reduced model) is measured at 1333.2. This suggests that (4525.4-1433.2)/4525.4 = 70.5% of the variability of the data can be explained by the fitted model which is quite a good fit considering the non-linearity of the system. However, optimal scale resolution levels were used to achieve this result as detailed in 5.11.3.

The results show that certain system variables are more significant than others when predicting the response. The results suggest that to adequately predict the damping of the 0.21Hz mode, the generation at generator buses 27,99,102 and 117 are required as these have the highest β parameters with low residual errors.

The ability of the GLM to produce significant predictors is further enhanced by the fact they partially match the average participation factors of the system. The PF column in table 5.5 shows that the highest β parameters are also the highest participation factors from the the dynamic model. This is the case for generator buses 27 and 99. However, GB102 and GB117 do not match as they are ranked differently in terms of regression

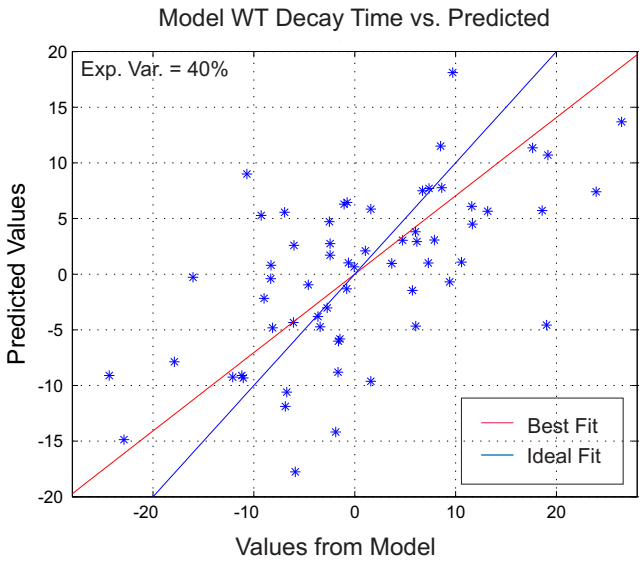
coefficients. GB52 has been included as it turns out to be a significant predictor when logic regression is applied to the system in chapter 6.

5.10 Reliability of GLM

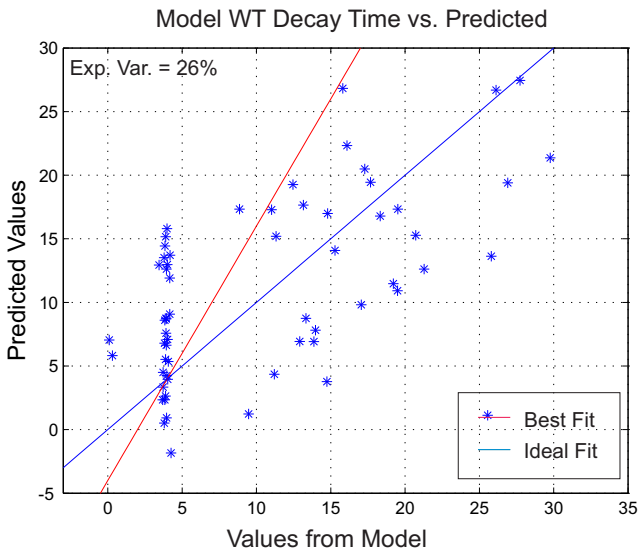
Figure 5.23 contains subplots that show the prediction capability of the GLM with (a) Wavelet coefficients and (b) raw values. The wavelet case has not been subject to variable selection or optimal resolution analysis. Using the Gaussian GLM model with the wavelet coefficients, allows a well fitted linear model to be developed. In 5.24(a) the actual wavelet coefficients from the model already display a quasi-linear relationship with predicted wavelet damping coefficients even though the data length is only 64 and variable selection hasn't been applied yet. The general linear model is essentially a linear regression with the model fitted with maximum likelihood and scored in terms of deviance instead of RMSE from a least squares fit. The blue best fit line represents a perfect fit and it is clear that the wavelet model is more linear than the raw model even though there is a lot of variance in the prediction.

This is in stark comparison to 5.24(b) where the raw data is ambiguous with no relationship between the predictor and response variables. It may be that the raw data requires a longer data length in order to improve the model fit but even at a data length of 64, it is not as good as the wavelet model. The model fit for the raw data can be measured in terms of deviances from the GLM model where the percentage of Y-Space variance explained can be measured. In this case it turns out that 26% of the model variance is explained by the predictor variables. In other words, the wavelet transformed data it much more suited to linear statistical models than raw data.

In this study, data lengths of 128 and 256 were used to test the wavelet transformed GLM's. These data points were derived from randomly determined loadflow scenarios and resulted in 128 and 256 sets of decay time constants for the 0.21Hz interarea mode as well as 128 and 256 sets of independent measurable system variables such as real and reactive power generation. Statistical analysis was performed on the active



(a) Linear Regression on Wavelet Coefficients



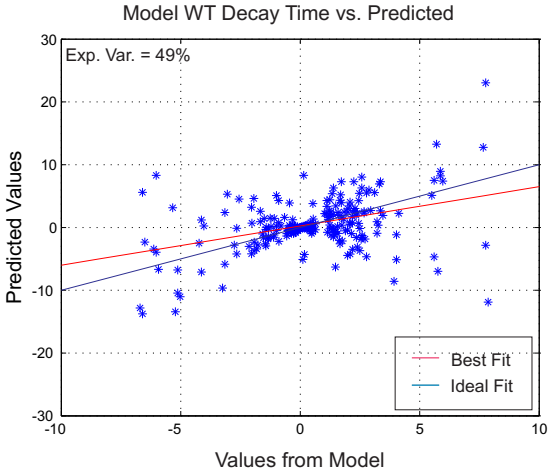
(b) Linear Regression on Raw Data

Figure 5.23: Actual Values versus Predicted Values derived from Linear Regression on Wavelet Coefficients and Raw Values

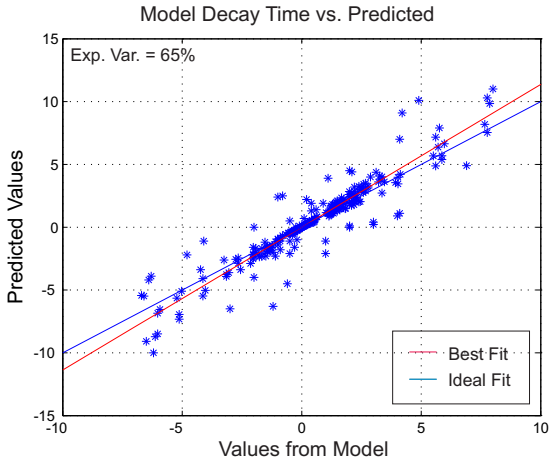
and reactive power generation variables as well as the decay time constant variables to assess if there was any relationship between decay time constant and active power. Both data lengths were used in the study to determine the size of dataset required to produce an acceptable result with minimum deviance.

The subfigures in figure 5.24 show the plots of the actual wavelet coefficients derived from the decay time constant versus the predicted values derived from the GLM. These plots are shown for various optimal combinations of wavelet resolution levels and data lengths. A best fit line has been added to the plots to show any linearity that would indicate a well fitted model if accompanied by low deviance. Subfigure 5.24(a) displays the GLM regression prediction when the whole frequency spectrum of the response and predictors are used i.e. every resolution level. The response and predictor variables in this case are of length 128. Referring to figure 5.25 it is clear that most of the dynamics of the signal occur at the second, third, fifth and sixth resolution levels.
It can be seen that when the full spectra is included in the GLM the prediction is quite poor with a lot of scatter as shown in subfigure 5.24(a). Here, the percentage of variance explained is 2722.1 with 127 degrees of freedom with the residual deviance at 1388.9 which corresponds to a percentage of variance explained of 49%.
In subfigure 5.24(b) a different scenario is presented. In this instance, only resolution levels 2,3,5 and 6 are used in the GLM as these resolution levels contain a large majority of the dynamics. Again, a data length of 128 is used. Resolution levels two, three, five and six correspond to sampling frequencies of the raw signal at $0.0625Hz$,$0.125Hz$, $0.5Hz$ and $1Hz$ Hz respectively. The wavelet coefficients determine what portion of the signal at that time is composed of frequency components at that resolution level i.e. level 5 coefficients determine how much of the raw signal is composed of the 0.5Hz dynamics at a particular point. It can be seen that when the redundant frequency bands of the signal are removed from the GLM the prediction is improved. In this case the null deviance is 2345.8 with 127 degrees of freedom with the residual deviance at 823.2. As a result 65% of the variation is explained in the model.

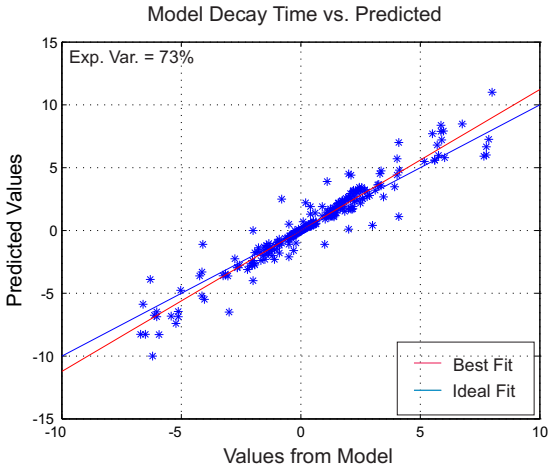
Subfigure 5.24(c) again shows a plot of the wavelet transformed actual decay time con-



(a) GLM Prediction with 128 data length and full wavelet spectra



(b) GLM Prediction with 128 data length and optimal scales



(c) GLM Prediction with 256 data length and optimal scales

Figure 5.24: Actual Wavelet Coefficients versus Predicted Wavelet Coefficients from power system model

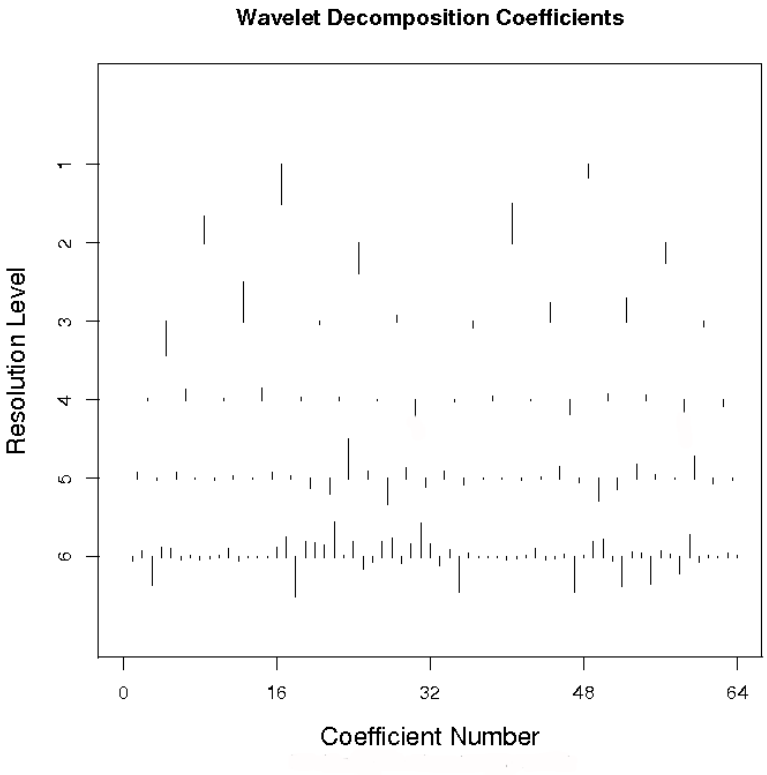


Figure 5.25: Wavelet Coefficients at various resolution levels derived from Icelandic 0.21Hz mode decay time constant for 128 length data.

stants versus the estimated decay time constant. In this case a data set of size 256 was used to see the effect the longer dataset would have on the wavelet transformed model. A longer data set allows more of the high resolution detail to be used in the GLM which means more datapoints. In this case, resolution levels 3,4,6 and 7 are used i.e. the same levels as before except doubled in length due to the longer window. Now, the mode explains 73% of the response variable variance which is a much more acceptable result. Once again, an improvement in the model is seen when the redundant data is removed and more detailed dynamic data included in the GLM.

The results show that a significant improvement can be achieved in the GLM if the optimal wavelet scale combinations are used in the analysis. Depending on the characteristics of the signal, longer data sets can be used to extract more details from the signal in terms of both temporal and frequency content. By determining which combinations of wavelet resolution levels reduce the overall deviance of the model it is possible to accurately determine the significant predictors that affect mode damping. At present, the optimal scales have not been automated to pick the best scoring models but this could be done by looping through each optimal scale mode i.e. levels 2,3,4 or 3,6,7 and determining the best model score.

In summary, a larger dataset allows a larger quantity of higher resolution data to be used in the analysis. Depending on the signal characteristics this may produce a more accurate wavelet transformed GLM as a greater length of wavelet scales can be combined to encompass all of the interesting dynamics in the system. However, due to the dynamic nature of the power system, using too long a dataset can result is some important information being lost.
As a consequence, the true predictors can become less meaningful due to the number of interactions taking place over the longer window. In general, less complex systems need shorter data sets to produce a good fit as the relationships are usually more obvious. More complex systems require longer data sets to establish a statistical relationship between the response and predictors. However, caution should be taken not to extend datasets over an unnecessarily long window as the model can become difficult to

interpret.

5.11 Icelandic GLM's with Reactive Power Variables

In the previous sections, general linear models were constructed with decay time constants and active power variables only. This was due to the fact that generators supplying active power to the network drive certain modes causing them to oscillate. It has already been shown that fairly good models can be developed using only active power as predictor variables.

In this section, reactive power has been added to the statistical models to see if there is any improvement in the model fit. Traditionally, system operators have looked to the active power conditions of the network when attempting to determine the causes of poorly damped events with reactive power providing supplemental information to the analysis as changes in generator power factor influences mode behaviour. The dynamic performance of a generator varies with the operating point of that generator. As a generators power factor is reduced, the contribution to damping may be reduced as well as an increase in reactive power tends to correlate with better damped electromechanical modes. As a result, it is useful to include reactive power flow in the model to give a more robust analysis of the mode dynamics.

Reducing the reactive power output tends to reduce the stability margin of the generator [2]. Since active and reactive power are linked via power factor, it is thought that by including reactive power in the statistical model more of the response variable variance can be explained. However, the inclusion of more system variables serves to add more dimensionality to the problem and requires more complex reasoning and thought as to causes of oscillatory dynamics.

5.11.1 Results from Icelandic GLM's with Reactive Power Variables

Figure 5.26 shows the actual wavelet transformed decay time constant variables (derived from the model) versus the predicted values for the GLM with reactive generation. Resolution levels 3,4,6 and 7 were used in the model development just as in the active

power case which corresponds to a data length of 216 (128+64+16+8). Pre-processing thresholding as well as variable selection was applied to both the active and reactive power components.

It can be seen how the model has been improved by the addition of the reactive power variables to the statistical model. The gradient of the best fit characteristic is 0.92 in comparison to 0.88 in the active power case which shows that the inclusion of the reactive power variables causes the model to approach a more linear state.
The null deviance of the model with the reactive power components is again 5233.3 with 207 degrees of freedom with residual deviance 1229.8. This corresponds to 5233.3-1229.8/5233.3 = 0.765 = 76.5% of the response variability being explained by the model. This compares to 73% derived from the optimal scale active power model used previously and shows that by including the reactive power components, the statistical model can be improved with increased prediction capability.

In the above example, the addition of the reactive power variables has served to improve the model fit by 3.5%. In this case, it may not be worth including reactive power in the model as the response variance can be explained almost as well with using active power variables alone. However, the model is improved and larger improvements can be obtained using different power system models. This allows insight to be gained into reactive power generation patterns that seem to be linked to poor damping performance. Additionally, if the reactive power variables contain little information on system damping then they can be easily removed from the statistical models allowing damping prediction exclusively with active power variables. For this reason, reactive power will be included in the analysis from now on. Table 5.6 displays the most sig-

Variable	Definition	Parameter Value	Residual Error	PF
G8	Gen Bus 27	0.0811	0.0143	1.000
G16	Gen Bus 99	0.0442	0.0111	0.8776
G17	Gen Bus 102	0.0361	7.4e-03	0.2113
G14(MV)	Gen Bus 32	-0.0107	1.6e-03	0.0023

Table 5.6: Parameters and Residual Error for Selected Variables from 256 point optimal scale GLM with Reactive Power Components

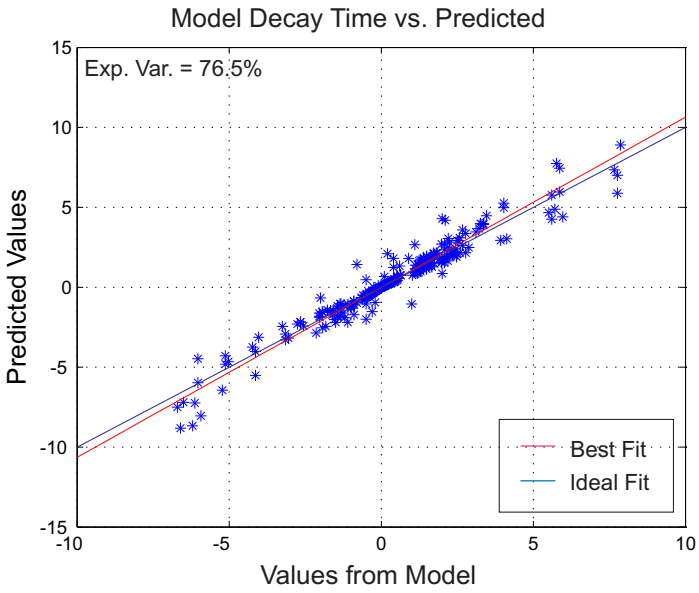


Figure 5.26: Actual wavelet transformed decay time constant from the model versus predicted values for 0.21Hz mode including reactive power variables. 256 data length with three highest resolution scales.

nificant coefficients derived from the general linear model as well as some of the most significant reactive power variables. It can be seen that the inclusion of the reactive power variables has served to alter the β coefficients but has not changed the order of the reported active power predictors. It is also evident that the reported predictors still agree with the participation factors calculated earlier. G8(GB27), G16(GB99) and G17(GB102) are still the most significant variables but in this case they have smaller residual errors due to the better fitted model. These improved predictors also perform better in the significance testing due to the improved model performance.

This is an interesting result as it shows that the inclusion of reactive power serves to improve the model while still retaining the rankings of the active power model. Figure 5.27 is a plot of normalized decay time constant superimposed with the normalized reactive power output at generator bus 32 which has generator G9 connected to it. It is noticeable that the reactive power output isn't that well correlated with the decay time constant profile. It is unlikely that GB32 reactive power is a major factor in this mode but when included in the model it serves to improve the model fit and it may furnish us with some information of the reactive power generation in the network that

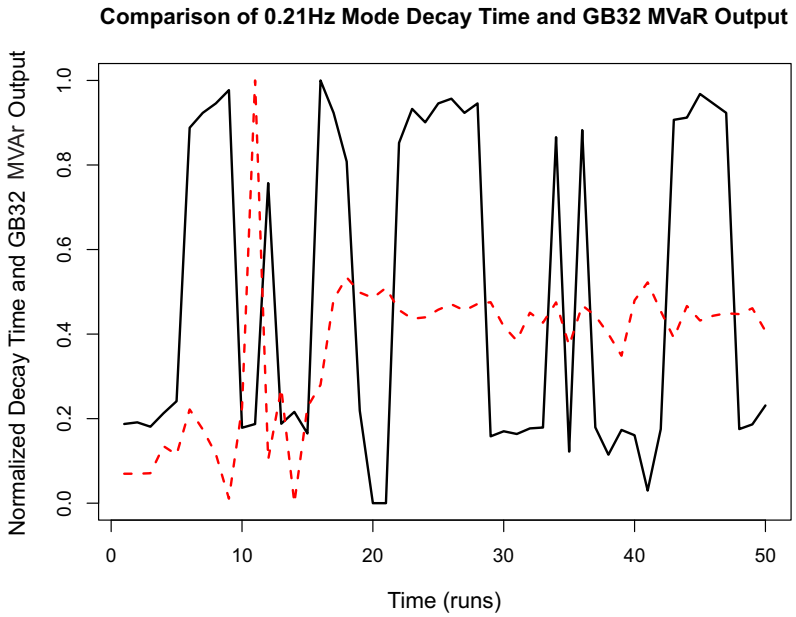


Figure 5.27: Normalized plot of 0.21Hz decay time constant and G9(GB32) MVAr Output.

seem to contribute to the poor decay time constants observed. Chapter 6 presents more work on the interaction effects of generator active and reactive power and how these interaction can cause modes to become poorly damped.

5.12 Sensitivity Analysis

5.12.1 Wavelet Transformed Sensitivity

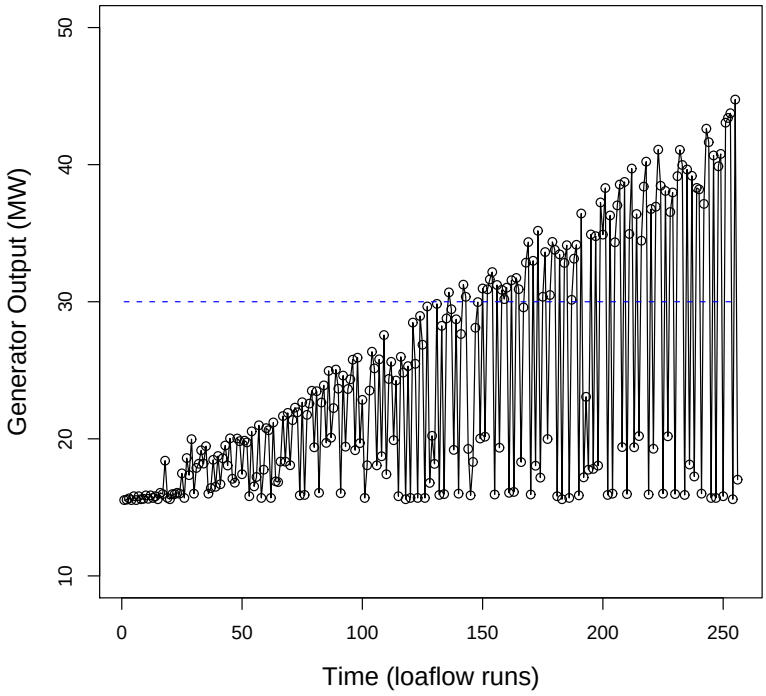
Since all the variables used in the model have been wavelet transformed, the sensitivity analysis required is less straightforward. The goal of the sensitivity analysis in this case is to show that as the high frequency wavelet coefficients of a significant generator increase, then so do the decay time constant high frequency wavelet coefficients. This is summarized in the following paragraph.

A large general linear regression coefficient suggests that a given predictor is strongly linked to the damping. Since the interesting dynamics in this case are relatively high frequency, we have chosen to use only these high frequency coefficients in our model as well as two lower frequency levels i.e. levels 3 and 4 in the 256 data length study.

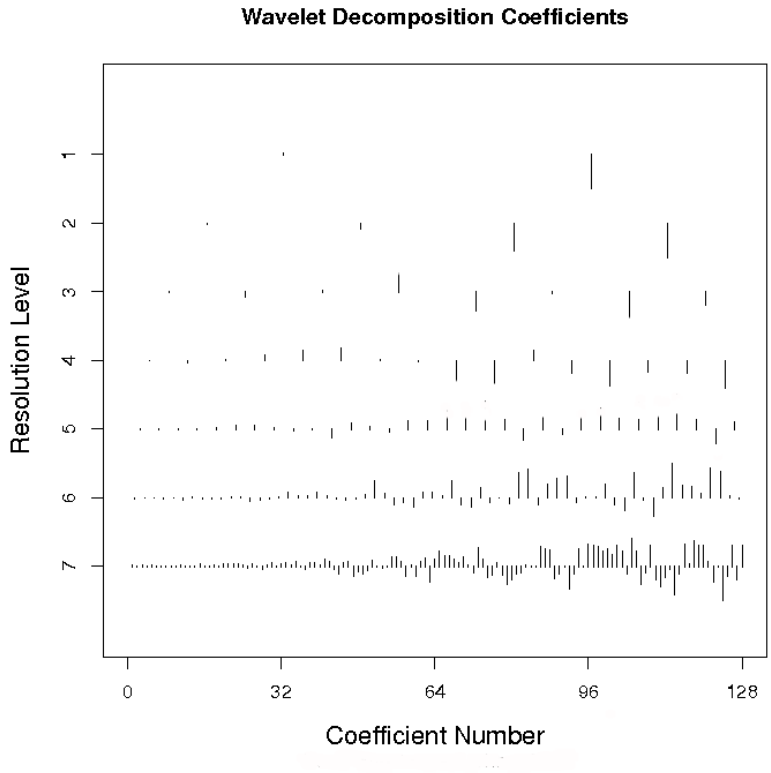
Therefore, an increase in these wavelet coeffients correspond to larger changes in the predictor variables at these frequencies. Thus, if we gradually increase the magnitude of the specified frequency changes in the generator output, we would expect to see corresponding increases in the decay time constant fluctuations i.e. larger decay time constant wavelet coefficients at these frequencies. Figure 5.28 contains subplots of (a) an oscillating ramp of generator output at GB27 and (b) the wavelet representation of the output. From subfigure 5.28(b) it can be seen how the wavelet coefficients at resolution levels 3,4,6 and 7 have gradually increased as the signal amplitude increases. This is due to the fact that the MW output increases in incrementally larger steps at the three specified frequencies corresponding to 0.0625Hz, 0.125Hz, 0.5Hz and 1Hz.

In figure 5.28(a), the GB27 machine has been ramped up from $\pm 50\%$ its nominal output, which is 30MW. However, instead of a smooth ramp up to full output, GB27 has been oscillating back and forth to the minimum limit of 15MW. This has been done in order to ramp the wavelet coefficients at the resolution levels used in the model. By doing this, it is possible to plot the resulting relationship between the wavelet transformed GB27 MW output and the wavelet transformed decay time constant. The gradients of the characteristics from the dynamic model and statistical model can then be compared to validate the GLM.

Figure 5.29 displays the raw decay time constant and its wavelet coefficients when GB27 has been subjected to an oscillating ramp while the rest of the system has been held as constant as possible. It is evident that as the GB27 output is ramped up, the 0.21Hz mode decay time constant gradually increases. The oscillatory input also has the effect of increasing the decay time constant wavelet coefficients as GB27 is increased. The increase in wavelet coefficients for the specified resolution levels is shown in subfigure 5.29(b). Here, the coefficients get gradually larger at levels 3,4,6 and 7 as the signal amplitude increases which shows there is a relationship between the GB27 generator and decay time constant frequency resolutions. The corresponding increase in resolution level 5 is relatively small as this was removed from the model as it was deemed redundant.

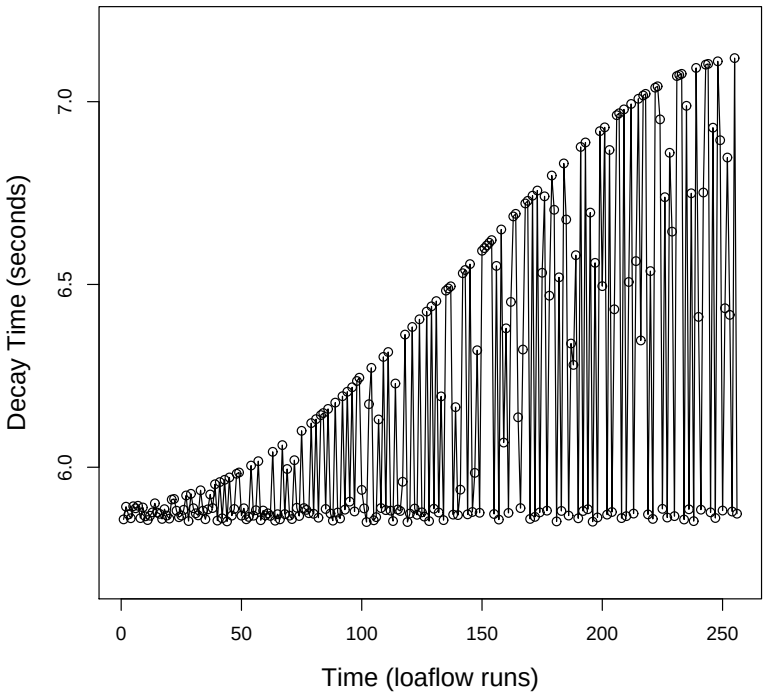


(a) Increasing MW Output at GB27

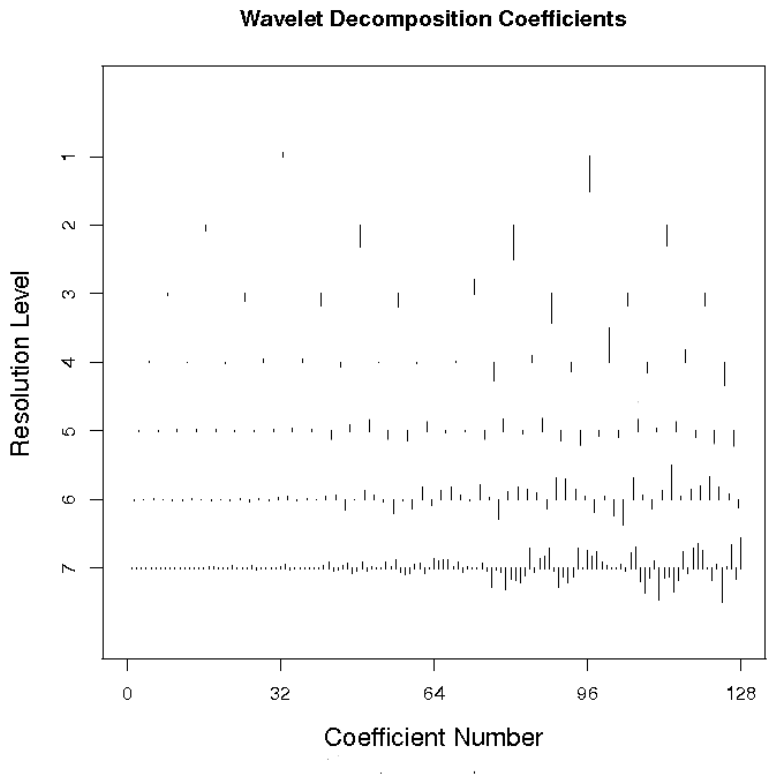


(b) Wavelet Coefficient Representation

Figure 5.28: GB27 Output subjected to an Oscillating Ramp



(a) Increasing MW Output at GB27



(b) Wavelet Coefficient Representation

Figure 5.29: 0.21Hz decay time constant Output

Figure 5.30 shows a plot of the decay time constant versus GB27 wavelet coefficients

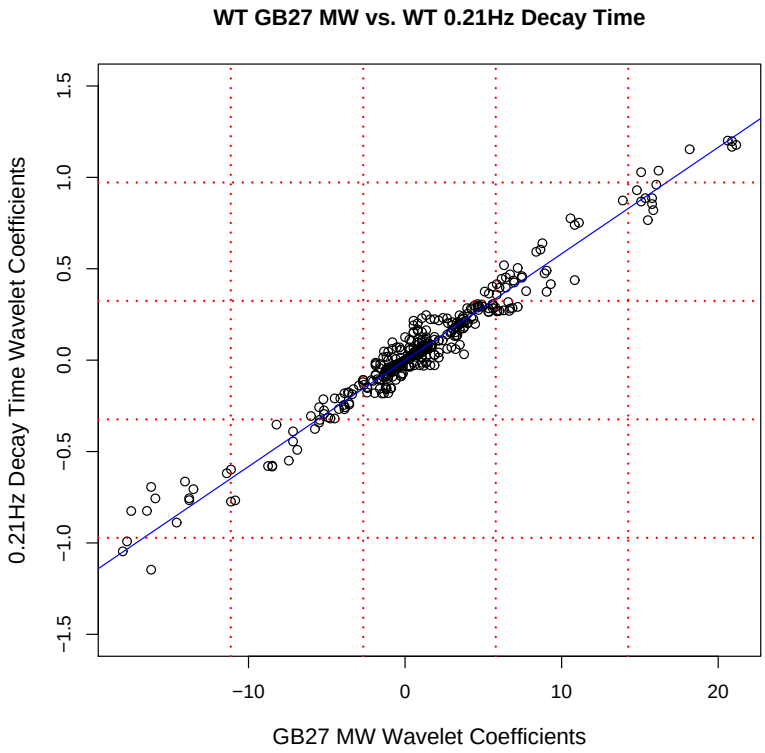


Figure 5.30: Wavelet Transformed GB27 MW vs. Wavelet Transformed decay time constant with buses held constant

derived from the dynamic model with GB27 MW output being ramped up, while keeping all other net bus injections as constant as possible. Resolution levels 3,4,6 and 7 have been used in this plot. The linearity of the wavelet coefficients show that the decay time constant is strongly related to the GB27 MW output. As the GB27 MW oscillation increases as shown in figure 5.28(a), so to do the oscillations in decay time constant, shown in figure 5.29(a). These oscillatory increases occur predominantly at resolution levels 3,4,6 and 7 as evidenced in figures 5.28(b) and 5.29(b). Figure 5.30 is closely linked to figure 5.10 as both plots show the relationship between GB27 and decay time constant. The difference here, is that figure 5.10 details larger decay time constant wavelet coefficients as the system is running as normal. However, figure 5.30 is derived from a system where only GB27 is modulated and as a result, the changes in decay time constant are smaller leading to smaller wavelet coefficients.

By using these sets of coefficients, a well fitted model has been developed that can determine the predictors strongly linked to damping. In the GB27 MW case, it can be seen that the wavelet coefficients of both GB27 and the decay time constant increase over time and since they occur over certain frequency bands, the wavelet transform can remove any redundant frequencies. In table 5.7, the gradients of the characteristics of

Variable	G8(GB27)	G16(GB99)	G17(GB102)
Dynamic Model	0.0672	0.0387	0.0265
Statistical Model	0.0834	0.0468	0.0331

Table 5.7: Comparison of 0.21Hz mode sensitivity to generator output from dynamic and statistical model

the 0.21Hz mode damping against generator outputs are compared for the dynamic and statistical models. It must be stressed that the gradients in question are not derived from the raw values used in statistical models i.e. raw decay time constant, raw active power etc. In this case, it is the wavelet coefficients of the response and predictor variables that are used in the model. These coefficients are then derived from the dynamic model, concatenated in a vector and plotted against one another (decay time constant vs. active power) to determine the resulting gradients (see figure 5.30). These are then presented in table 5.7 for comparison.

From the plots of MW wavelet coefficients vs. decay time constant coefficients, the best linear characteristic is plotted and the gradient is noted in table 5.7. The gradient from the statistical model is the relevant β parameters derived from the GLM.
It was noted that the generator outputs that were selected as predictors show a very clear relationship between damping and generator output as illustrated in figures 5.31, 5.32 and 5.33. Also, a good match is achieved between the gradient of the characteristics derived from the statistical model and the dynamic model as shown in table 5.7. This confirms that the wavelet coefficients can be used to determine significant predictors from the model even when decay time constants are large and the system is close to its stability limits. In the 0.21Hz case, the maximum decay time constant was recorded at 22 seconds which corresponds to a damping ratio of 0.0018 which is well inside the

non-linear range of mode damping.

This example shows how optimal scale wavelet coefficients can be used even when the mode is close to the stability limits. Since the wavelet transform causes the relationship between the response and predictors to become more linear at the stability limits, it allows the data to be modelled in a Gaussian GLM model. As a result, the sensitivities of the model will be highlighted in the GLM even when a very resonant mode with a low damping ratio is detected.

5.12.2 Statistical Model versus Dynamic Model

This section presents a selection of results from a sensitivity analysis that was conducted to determine how the 0.21Hz system mode decay time constant reacted to changes in the significant predictors obtained from the wavelet transformed GLM (tables 5.5 and 5.6).

Figure 5.31 shows the sensitivity of the 0.21Hz mode to the active power output at generator bus 27 in the Icelandic model. GB27 nominally outputs 30MW and the graph shows the effect ramping this generator up from 0.5*30 = 15MW to 1.5*30 = 45MW has on the mode decay time constant. Although the generator is smaller than others reported, it clearly has an effect on damping. This phenomenon is explored more in chapter 6 which studies the effects of smaller machines interacting with larger ones to produce long decay time constants.

It can be seen how the mode becomes more resonant as the generator output is ramped up between factors of 0.5 and 1.5 times its nominal value. It is also interesting to note how the decay time constant levels around 7.1 seconds compared to the maximum decay time constant of 22 seconds recorded during some of the events. This result shows that GB27 has an influence on the mode damping and that it may indeed act as a trigger for instability. However, it is not the sole reason for instability as evidenced by the short decay time constants observed. It must also interact with a host of other factors in order to produce the higher decay time constants observed in the model.

Figure 5.32 also shows the sensitivity of the mode to the active power of the gener-

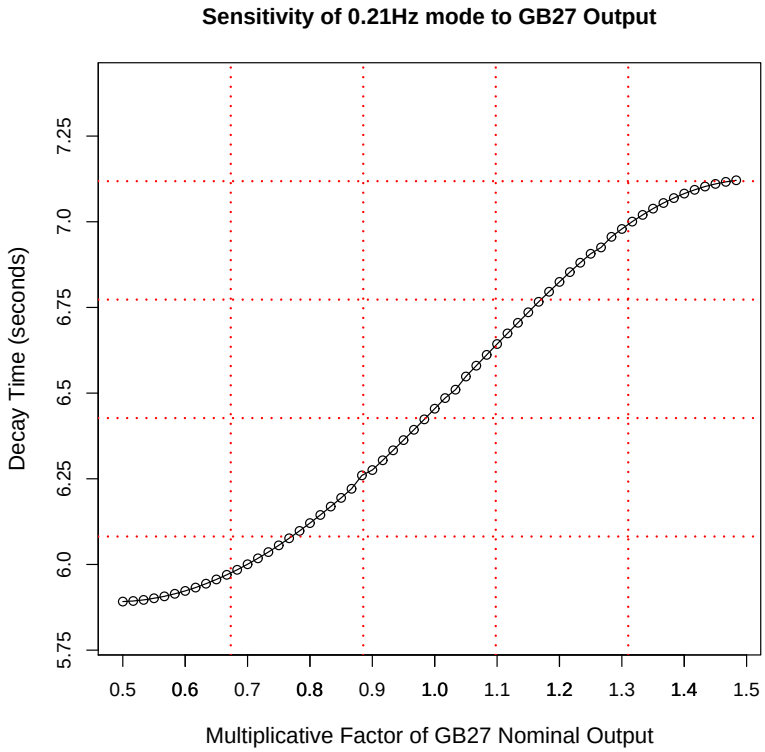


Figure 5.31: Sensitivity of 0.21Hz mode to GB27 active power output

ator at bus GB99. A generator with a nominal 50MW output is connected to GB99. Although the change in damping is not as smooth and logistic as GB27 a similar relationship can be seen. It is also interesting to note how the total change in the mode decay time constant is greater than the GB27 case although it has a smaller gradient. This is due to GB99 being a much larger machine although decay time constant is much more sensitive to the GB27 output as evidenced by the β coefficients. This shows how different machines participate to varying degrees in mode dynamics.

Figure 5.33 displays the decay time constant record with the active generator at generator bus 102 ramped from 0.5 to 1.5 nominal output. GB102 is connected to a 60MW generator in the model. The relationship in this case appears to be more linear as opposed to the logistic relationship of GB27 and again the decay time constant change is minimal, only changing by approximately 0.5 seconds. However, these examples do

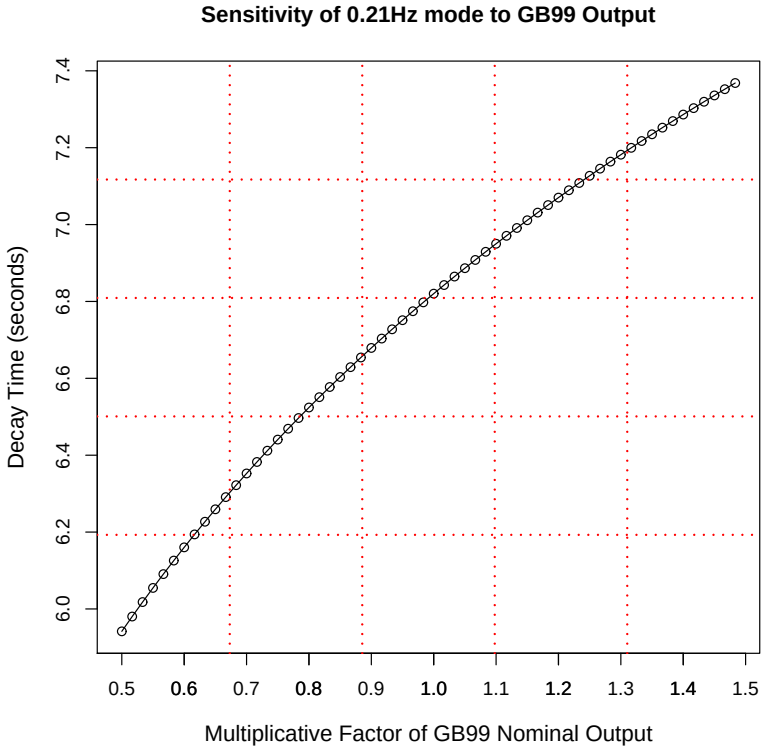


Figure 5.32: Sensitivity of 0.21Hz mode to GB99 active power output

show that the GLM is capable of extracting significant predictors that are critical in predicting the system response. These strongly linked predictors may also provide a means of controlling mode damping before, during or after poorly damped events.

Figure 5.34 displays two subplots of the most significant generators derived from the GLM for the 0.21Hz Icelandic mode. The plots show a section of the decay time constant record compared to a section of the generator outputs for the 256 data point analysis. A smaller section was used so that the comparison can be seen as longer plots tend to clutter any obvious relationships.

It can be seen that both MW outputs are quite well correlated with the periods of high decay time constants. In fact, between GB27 and GB99 almost all of the instability has been explained. This further confirms the fact that the wavelet transformed GLM is able to determine significant predictors that can be used to control mode damping. However, from previous studies we know that dynamic events are rarely a case of one way interactions between a single predictor and damping.

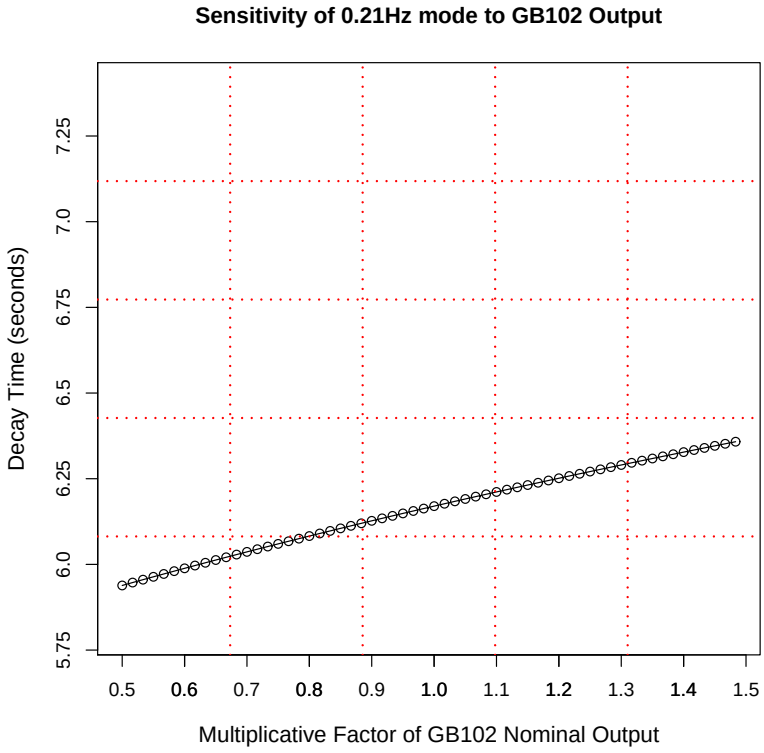
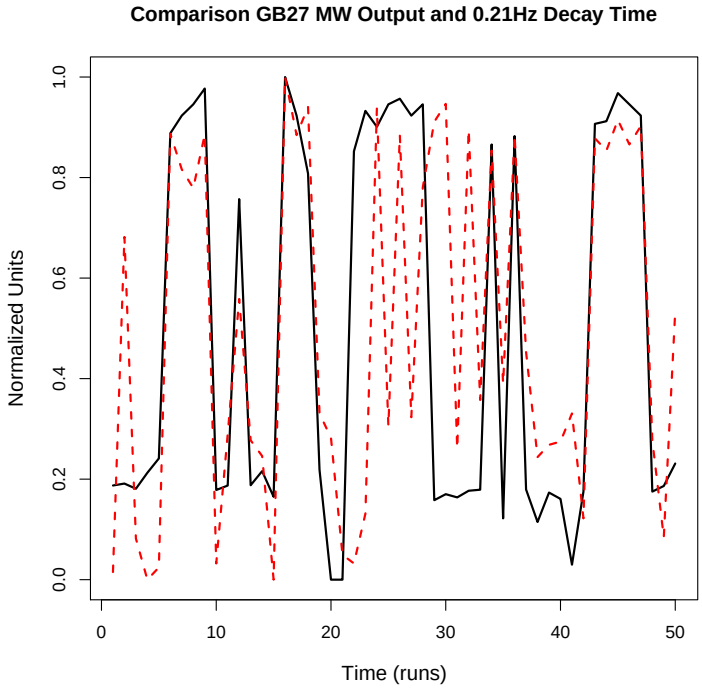


Figure 5.33: Sensitivity of 0.21Hz mode to GB102 active power output

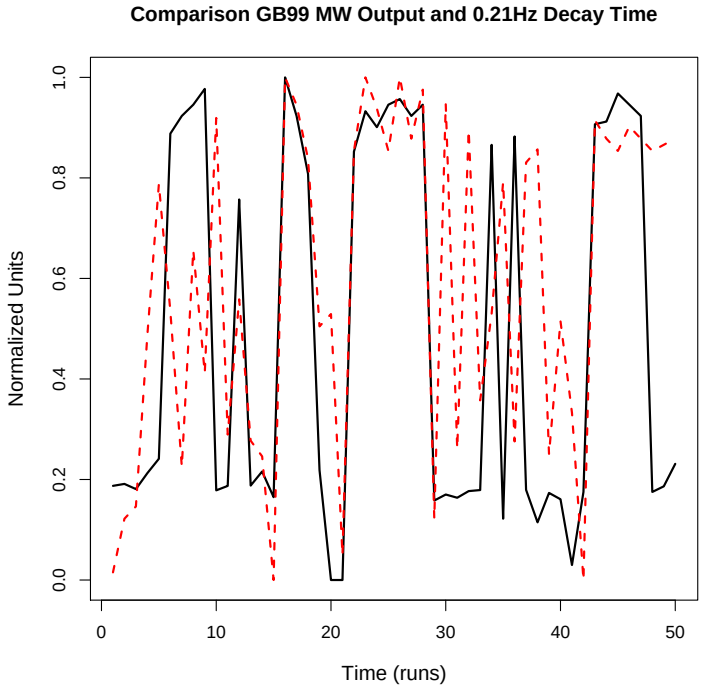
In the majority of cases the problem is much more complex and depends on a host of interactions. The WGLM is a powerful enough method to extract one or two way interactions at most. However, the ultimate goal is to model the more complex interactions so that a more robust damping control method can be established. It was noted that the generator outputs that were selected as predictors show a very clear relationship between damping and generator output as illustrated in figures 5.31, 5.32 and 5.33 as well as figure 5.34. Chapter 6 deals with possible interactions between these variables that tend to make poor damping more likely as opposed to single one-way way interactions.

5.12.3 Icelandic GLM's with Variable Selection

In section 5.6.3 we discussed how some preliminary thresholding can be done on the system state variables to simplify the resulting GLM. Table 5.8 contains information on four of the variables removed in the preliminary stage. Both the mean and the variance



(a) 0.21Hz decay time constant plotted with GB27 MW Output



(b) 0.21Hz decay time constant plotted with GB99 MW Output

Figure 5.34: Comparison of Nominal 0.21Hz decay time constants and G8(GB27) and G16(GB99) MW Outputs

has been shown for the 128 dataset model. It can be seen that both the means and the variance are relatively low with the variances corresponding to maximum swings of 1 to 1.5MW from the mean (see figure 5.8). Since these generators are supplying negligible oscillation energy to the mode in question and because they do not swing enough to seriously perturb the system they have been omitted from the model. In

Variable	Definition	Mean	Variance
19	Gen Bus 50	6.04	0.43
23	Gen Bus 98	4.21	0.44
32	Gen Bus 142	3.06	0.27
34	Gen Bus 186	6.11	0.47

Table 5.8: Mean and Variance values for predictors removed in preliminary thresholding

section 5.3.2 we discussed two methods of variable selection that allowed us to remove insignificant and multicollinear variables from the statistical models. In this section both mutual information and PLSR will be applied to the Icelandic data to determine if any redundant variables can be excluded.

Figure 5.35 shows the coefficients derived form a mutual information model and PLSR model for the Icelandic system with 256 data points. The green bars indicate redundant variables that are excluded from the model. Variables 2,12 and 15 have been excluded as they are deemed to have negligible mutual information with the response signal (decay time constant). Variables 6,10,15,16 and 26 have been excluded as the PLSR model has deemed them to be insignificant because the remaining variables that form the significant latent vectors explain the greatest proportion of the response variance. Variable 15 is rejected by both PLSR and MI. The total redundant variables from both PLSR and MI relate to generator buses 3,13,19,28,33,36 and 126. As a result, the variables in question cannot be the true underlying cause of response variability. In the case of the redundant predictors variables from PLSR, they may be correlated with a true predictor (giving them a high MI coefficient) but they are not the cause of poor damping as they do not account for much Y-space variation. However, they may be correlated with significant predictors and can be added after the analysis for cases like generators sitting in the same plant etc.

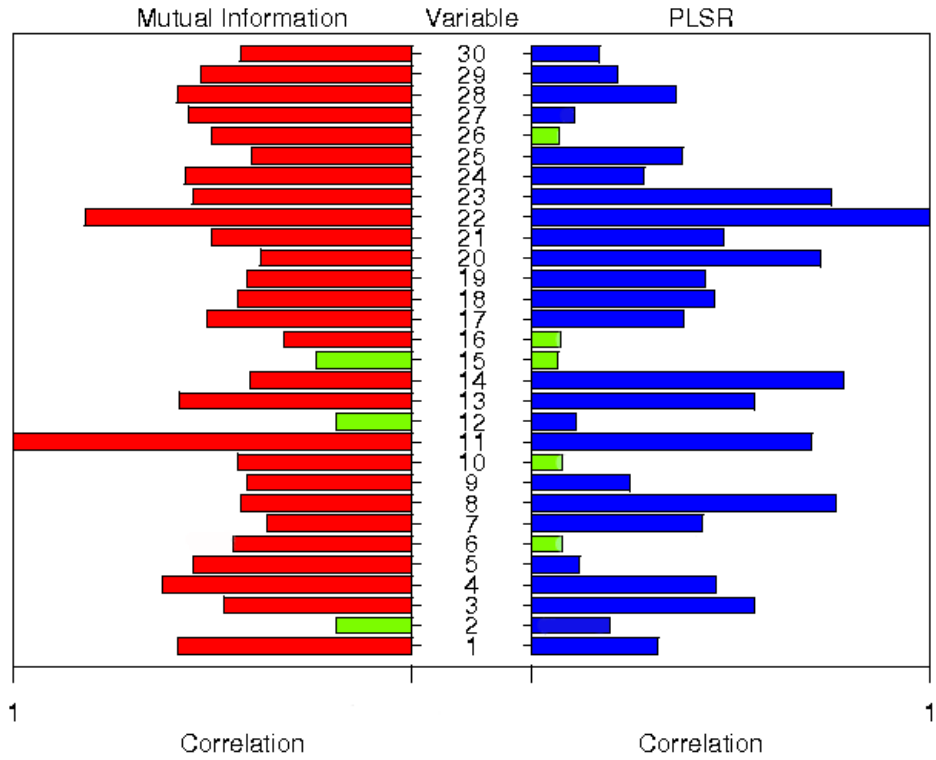


Figure 5.35: Pyramid plot showing coefficients derived from mutual information and PLSR. Green bars indicate redundant variables.

Although both the mutual information and PLSR don't always reconcile, it is useful to combine them to remove redundancy from the model. The MI gives us an idea of information shared by two signals whereas the PLSR allows us to determine which variables have a negligible effect on the response in terms of their interaction with other predictors i.e. the PLSR latent vector. However, care has to be taken to ensure too many variables aren't removed as this could inhibit the model prediction capability. Figure 5.36 shows a plot of actual wavelet coeffieicents versus predicted coefficients

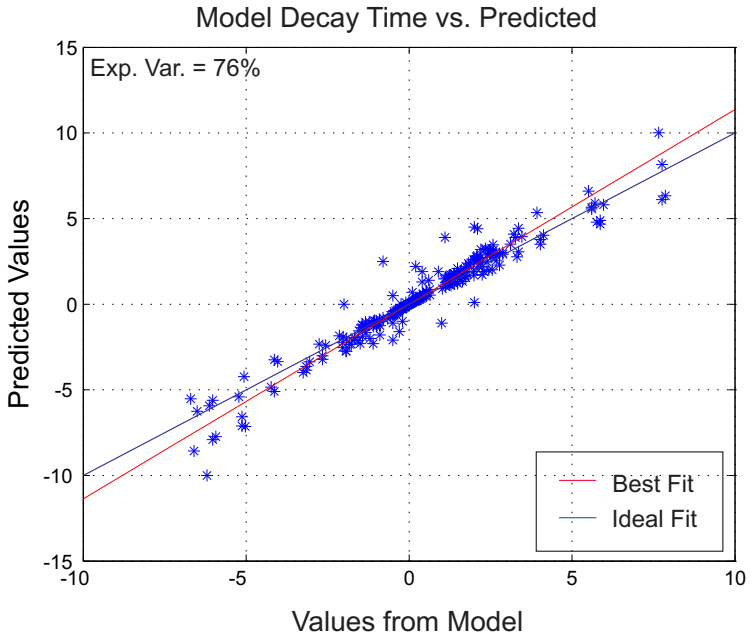


Figure 5.36: Actual wavelet coefficients from the model versus predicted wavelet coefficients for variable selection model

for the Icelandic model with 256 data points. In this case the third, fourth, sixth and seventh resolution levels are used again and variable selection has been utilized with variables 2,6,10,12,15,16 and 26 (generators at buses 3,13,19,28,33,36,126) being removed from the model via the MI/PLSR step as well as variables 19,23,32 and 34 (GB 50,98,142 and 186) from preliminary thresholding. This model has been developed using only active power as predictor variables.

It can be seen that the model is now a slightly better fit than the same case without

variable selection (see figure 5.24(c)). The null deviance of the model is 5154.5 with 207 degrees of freedom with the residual deviance measuring 1226.4 on 96 degrees of freedom. These results mean that 76% of the response variable variability is explained by the GLM. This result is better that the non-variable selection active power case shown in figure 5.24(c). This model could be used in a source location study with the strongly linked predictors being used to control damping, subject to residual error and significance tests.

It is interesting to note that the variable selection active power model is almost as good as the model with the reactive power included. With the use of reactive power with no variable selection the model produced a 76.5% fit. With variable selection on the active power variables the model was able to produce a 76% accuracy.
Combining these two procedures, results in a model with a 78.2% fit which shows that by combining variable selection and reactive power, a better model can be obtained. By increasing the data length of the model, it is thought that an even better model score could be achieved. However, shorter data sets allow higher resolution events to me studied which may be of more importance. It must be noted that the reactive power serves to reduce the model error while still maintaining the rank of the significant variables in most cases. This is covered in more detail in chapters 6 and 7 where logic regression is also optimized using variable selection and reactive power variables.

The best model score presented in this chapter (78.2% with MW and MVAr generation predictors), is less than the value presented in [45], where a model with a root mean squared error (RMSE) of 0.11 was achieved. This meant that 89% of the dependent variable variance was explained by the predictor set using a linear multiple regression. However, linear regression was applied to the 0.21Hz Icelandic mode from section 5.9 and it was found that no reasonable prediction could be made as shown in figure 5.23(b). In addition, the variable selction techniques were also applied to both the 0.21Hz Icelandic case and 0.43Hz 16 Machine mode and the resulting prediction was still found to be unsatisfactory.
The study in [45] was confined to a small part of the system where the predictors used

were known to affect the 0.5Hz mode due to their proximity and historical trends. In the case of the Icelandic and 16 machine models, no such information was available and as such, assumptions based on the proximity of generators couldn't be made as there was no historical data.

The results in [45] are produced from multiple linear regression on the power system data. Due to the fact that power system data is not normally distributed and because its generally heteroscedastic (increasing variance with magnitude) the model fit may be overestimated as is common when this is the case.

The prediction from the WGLM shown in figure 5.36 was able to predict 76% of the y variable variance with only active power as the predictor variable. In figure 5.26, 76.5% was achieved with both active and reactive power variables. This is compared to 89% in [45] where active and reactive generation was used as well as line flows, bus voltages and angles. By including more variety of system states in the WGLM's a better fitted model could have been achieved.

The results in figures 5.26 and 5.36 are derived from 256 length data. The scores obtained of 76.5% and 76% are roughly the same for the testing set in [45] when a 250 length dataset is used. The results presented in this chapter were derived from decay time constant data close to the stability limits of the system where the relationship between the active and reactive power variables is non-linear. This, combined with the fact that the system data isn't normally distributed, makes multiple linear regression unsuitable for power system analysis on raw data. This was also found to be the case for the 16 machine model (figure 5.14) and a number of other models used during this research.

5.13 Summary

Using a 16 Machine Model as well as an Icelandic System model, this chapter presented results which supported the use of wavelet transformed general linear models (WGLM's) in determining system variables that can be used to predict and control mode dynamics.

The results showed that by wavelet transforming both the response and predictor variables, GLM's can be developed that can accurately predict the damping close to the system stability limits. This transformation caused the relationship between the dependent (damping) and independent variables (active/reactive power) to become more linear which meant they could be used in a GLM. In addition, the distribution of the decay time constant profiles became near-normal in contrast to the raw decay time constants which exhibited an ambiguous distribution. This further increased the applicability of the transformed datasets to GLM's.

In the 16 machine case, optimal frequency bands were extracted for data lengths of 128 and it was found that the general linear model could explain 71% of the variance in the transformed decay time constant. This was a substantial improvement on the raw case where only 28% of the variance was explained. Variable selection in the form of mutual information (MI) and partial least squares regression (PLSR) were then applied to the model and it was found that 75% variance could be accounted for.

The Icelandic 0.21Hz mode was then used with the loadflow data to develop a GLM. Initially, 49% of the response variance was explained for the full frequency spectrum which was increased to 73% with optimal frequency scale application. By including reactive power in the model, more statistical information was made available and as such the model deviance was reduced and 76.5% of the response variable was explained. By incorporating both the reactive power and variable selection, the best scoring model was achieved with a 78.2% variance accounted for.
Even though the reactive power only slightly improves the model, it will be included in subsequent studies as information on the relationship between mode stability and reactive generation could be used for more robust control of system stability. In the Icelandic case the reactive power was of little value but this isn't a general conclusion and it has been seen in other studies [2] how reactive power can significantly contribute to the statistical model.

The Icelandic 0.21Hz model results were contrasted with the linear regression model in

[45] and although the latter produced a better fit, it was shown that the same approach couldn't produce the same result in the well validated Icelandic and 16 machine models. In addition, the model in [45] was developed with a broader range of system variables that were in close proximity to the 0.5Hz mode and were known to have an effect on it. By combining optimal scales with variable selection and reactive power data, a more superior model can be developed to that of the 0.21Hz linear regression mode, which will produce much more accurate predictors that can be used to control damping.

Sensitivity analysis was performed on the most significant predictors derived from the GLM and it was shown that they had an effect on the 0.21Hz mode damping. However, the magnitude of the decay time constant increases when significant individual generators were ramped up (while the rest of the system was held constant), were much smaller than the decay time constant increases during poorly damped events when system wasn't constrained. The dynamic model was reconciled with the statistical model as the dynamic gradients and statistical regression coefficients had similar magnitudes which further validated the use of the WGLM in extracting significant active and reactive power variables that could be used to control the mode decay time constant and therefore system stability.

Chapter 6

Investigating Events using Logic Regression

6.1 Introduction

Various attempts at solving the source location problem have been detailed in the literature [2] [45] and each have used some form of variable selection with a linear regression model. The dynamic behaviour of a power system is such that periods of instability may not solely arise due to large generators sitting beside each other in large plants. While these large machines operating at near full capacity certainly have an effect on modes in the system, the triggering action may be something more inconspicuous like a smaller generator or load that has the effect of exciting certain system modes.

These interaction effects in the system are closely related to *sensitivity analysis* [77] which aims to find out how the variation in output is caused by sources of variation in the input. Currently, it is possible to calculate various participation factors that determine the contribution of various generators to specific system modes. These participation factors, though useful, only describe a one-way interaction and do not account for some of the dynamics witnessed in the systems as in many cases, the dynamics are caused by an interaction of generators. Participation factors are often used in *root cause analysis* [78] that attempts to determine the fundamental triggers for undesirable system behaviour.

As previously stated, many of the statistical tools used to locate the sources of oscillation are not fully adequate as they can only deal with the combination of one or two variables at once. The non-linearity of the system does not lend itself to linear statistical techniques where operation near stability limits means that some statistical information is lost. Instead, a more powerful technique is required that can explain how combinations of system variables interact to cause the poor damping observed in the system.

Regression is arguably the most important tool in the field of statistics as it is used to analyze data and to infer relationships between a set of predictors and response variables. However, in most regression problems a model is developed that only relates the main effects i.e. singular predictors to the response. Although interactions between predictors are sometimes considered as well, those interactions are usually kept very simple (one or two-way interactions at most).

It is often the case that a change in response is caused by the interaction of many predictors. For example, in [79] the authors are concerned about the relationship between disease loci and complex traits. A key emphasis is placed on the interaction of disease loci and potential shortcomings of methods that do not take those interactions properly into account. This problem is analogous to the electrical power system where the interaction of many different components can produce poorly damped events.

These problems illustrate the importance of a robust method in dealing with the interactions of system components. It can be stated as follows; given a set of predictors X, how can we create new, better predictors for the response by considering the combinations of those predictors. As it turns out, this problem can be addressed by converting all the response and predictor variables into binary values of 1 or 0. When the variables are in boolean format the goal now is to determine decision rules such as "if X_1,X_4,X_5 *and* X_9" are true, or "X_5 *or* X_6 but *not* X_8", then the response variable is more likely to be low i.e. belong to boolean class 0. In other words, we find Boolean statements

involving binary predictors that enhance the prediction of the response.
In actual fact, it is not necessary for the response variable to be a binary value as logic models can still be developed. This will be explained later in this chapter.

Appendix B contains an appreciable introduction to Logic Regression with an extensive list of the laws and symbols used. Information on model fitting procedures and cross-validation methods are also included as an aid to understanding some of the logic regression concepts. The key point of this chapter is to show how different combinations of system variables can be scored in terms of their effect on damping and how these combinations compare to the participation factors derived from state-space analysis.

6.2 Logic Regression in the Power System

As discussed in previous chapters, one of the main goals of this work is to study the effects of generator interactions on the power system dynamics. In chapter 5 section 5.12.2, it was shown how the most significant generator derived from the wavelet transformed GLM (GB27) caused a relatively small increase in the decay time constant of the Icelandic 0.21Hz mode to approximately 7.1 seconds. When compared to the maximum decay time constant of 22 seconds recorded during the events used in the study, it was clear that GB27 wasn't solely responsible for the poor damping.

In this case it is perfectly reasonable to assume that a number of factors are working together to create the conditions for these relatively long decay time constants. In the previous chapters, we focused mainly on active power variables as these are usually the strongest predictors in relation to mode damping.
In this chapter, the interactions between machines will be studied using their MW and MVAr outputs to determine if there are certain combinations of generators that make poor damping more likely. In some cases, one machine in the system can be used to successfully control damping across a range of system conditions. However, even in these cases the generator only exerts its effect because other generators are participating at outputs that are conducive to poor damping.

It may be the case that if the initial generator is held constant and a seemingly less significant generator is modulated, it too could have a major effect on damping. By establishing these generator interactions, a more robust control of mode damping can be acquired as all the interacting generators, that have the potential to cause poor damping, can be controlled to ensure the system remains stable.
This approach gives an unprecedented amount of control in contrast to the simpler one way interactions of normal regression where only single machines are reported as possible controlling factors. It also presents itself as a more economically viable solution as a range of generators can be used to control damping instead of a single generator, which gives the operator a number of possible solutions.

6.2.1 Using Dichotomous Variables

Traditional methods used to study the source location problem have always involved some sort of linear regression analysis. By using the power system data in a linear regression, the aim is to produce a well fitted model that has good response variable prediction capability. Good prediction capability of a model indicates that the β coefficients in the expression are accurate and can then be used to modify the response variable. The dependent and independent variables may also be transformed to produce more linear relationships, or to control the variance of the response, both of which make the data more suitable for linear regression. This was covered in chapter 5 where wavelets were used to linearize the relationship between the response and predictor variables.

In the logic regression case with a binomial link function i.e. using predictor combinations to predict log of the odds of excessive decay time constants (see section 5.4.3), the response variable is no longer modelled in terms of its raw values. In section 5.4.3 binomial regression was introduced and this type of generalized linear model is closely related to logic regression. In binomial regression, the response variable is no longer modelled as a raw value. Instead, the probability of an event happening is introduced, in our case, the probability of damping being poor or not poor.

The binomial case went a step further by using a link function to model non-linear relationships between the predictors and the *probability* of an event occurring i.e. probability of getting a 1 or poor damping.

The canonical function of the binomial distribution, called the *logit*, was discussed in the previous chapter and is shown in figure 6.1. This function models a logistic relation-

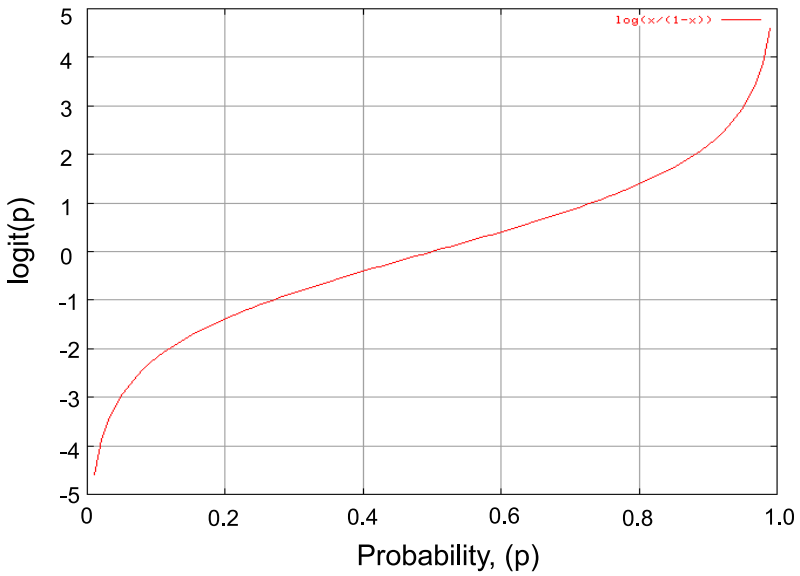


Figure 6.1: Logit Link Function

ship between the probability of an event and the raw predictors in the model. This has interesting consequences in terms of the source location problem as the dichotomous outcome of the model is a 1 or a 0 which is analogous to the event detection step i.e. thresholding a damping record to produce a 0 (good damping - no alarm) and a 1 (poor damping - trigger alarm). Thus, the response data can be thresholded to produce a stream of binary data that can be used in a binomial regression model.

This dichotomous representation is also the foundation for logic regression. In the linear logic regression case the response variable is left in its raw format while the predictors are transformed into binary values. Again, this allows the interactions in the system to be highlighted which better explains the variation in the response.

By dichotomizing the predictor and/or response variable(s) it makes it possible to find

Boolean statements, involving the binary predictors, that enhance the prediction of the response.

In more specific terms: Let X_1,...,X_k be raw/binary predictors, and let Y be a response variable. The aim of logic regression is to fit regression models of the form g(E[Y]) = b_0 + b_1L_1+ ...+ b_n L_n, where L_n is a Boolean expression of the predictors X, such as L_n=[(X_2 or X_4) and X_7]. The above framework includes many forms of regression, such as linear regression (g(E[Y])=E[Y]) and logistic regression (g(E[Y])=log(E[Y]/(1-E[Y]))). For every model type, a score function is defined that reflects the "quality" of the model under consideration.

In the following sections, the logic regression methodology will be applied to both wavelet transformed and raw data to determine the interactions that make poor damping more likely. These interactions will be thoroughly tested against participation factors, mutual information and the GLM's to determine if the results can be validated. A sensitivity analysis will also be performed on the reported interactions to gauge the accuracy of the method.

6.3 Developing a Logic Model

The development of a logic model is quite similar to the development of a general linear model, which was covered in chapter 5. In this chapter, the 0.21Hz Icelandic model mode will be used to demonstrate the logic regression step before it is used in the case studies of chapter 7. The active and reactive power variables used in this chapter will be subject to pre-processing, thresholding and variable selection as described in chapter 5. The remaining datasets will then be used to develop the logic model. The first step in this process is thresholding which transforms the variables into binary values.

6.3.1 Thresholding

In order to setup the binary values required for logic regression, it is first necessary to use a thresholding method so that the power system signals are correctly represented. This is a difficult problem due to the dynamic nature of the network and the fact that

the logic models are sensitive to the thresholding values.
In terms of the response variable (decay time constant) it is relatively easy to determine a thresholding value as engineers are generally in agreement over values of decay time constant that pose a threat to system stability. In these cases, alarms can be set to be triggered at decay time constants ranging from 5 seconds to 20 seconds. This threshold is termed "engineering threshold" in figure 6.2 and throughout the rest of this chapter.

In the Icelandic 0.21Hz mode case the threshold was initially set at 8.2 seconds as this happened to be the mean value of the test dataset. Although the mean value is usually a good threshold, the thresholding of the response variables, as well as the predictors, depends on the actual signal characteristics. Specifically, the magnitude of the dynamic changes in the signal and the length of time a signal is above the lower bound alarm threshold, which was set at 5 seconds for this study.
This stems from the fact that some power system data signals may have a relatively low variance and as a consequence, a mean value threshold setting would result in half the values in the decay time constant record being assigned 1 and the other half being assigned 0. Of course, this would only happen if the mean value of the dataset was above the initial threshold. If this were not true for the damping, then a number of the values above the mean threshold would still be below the 5 second lower bound threshold and would thus be assigned a binary 0.

Figure 6.2 shows a plot of a decay time constant signal with the mean value (mean taken over all the samples in the window) and an engineering threshold (determined by engineering judgment from past experience). In the case of the low variance signal shown in figure 6.2, relatively small decay time constants will be assigned high values i.e. 1. If a large proportion of the decay time constant binary profiles are 1's then this could lead to the model assigning significance to interactions that produce relatively small decay time constants as opposed to some of the more extreme values. This means that the reported predictors will in fact be describing larger sections of the decay time constant window, most of which are sufficiently damped.
However, if a threshold is chosen so only the more extreme decay time constants are

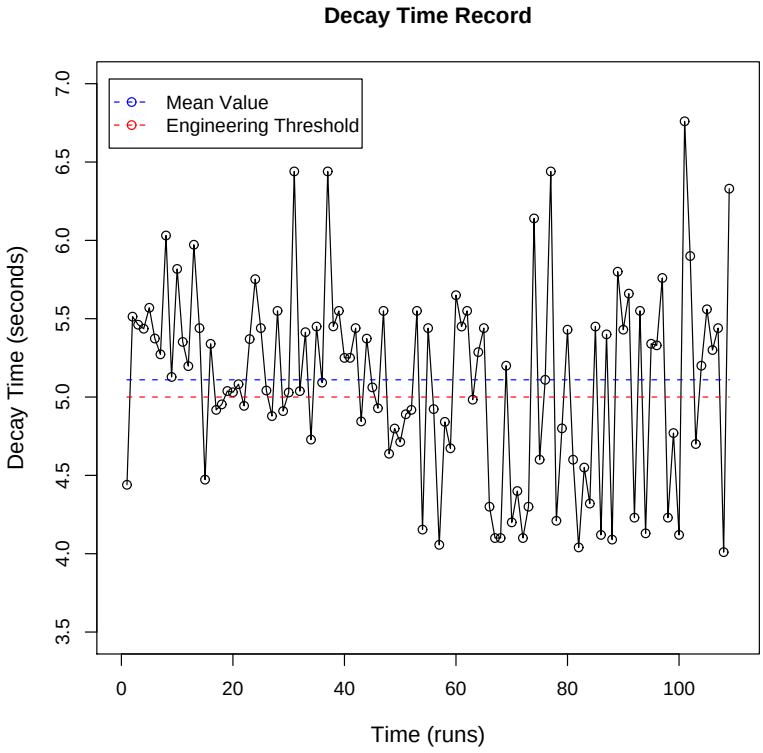


Figure 6.2: Low Variance decay time constant record showing engineering threshold and decay time constant mean

assigned a 1 then it is possible to extract the interactions that are responsible for poor damping.

A signal with a high variance may suffer from a similar problem. In the power system, a high variance signal may still have a low mean if a relatively small proportion of the signal is spent at the signal maximum. In this case, if the mean value is used then a large number of high values (1's) will be assigned and sensitivities to the extreme damping values will be smudged along with the smaller decay time constant values. This is illustrated in figure 6.3 when the mean value of the signal is sufficiently low to cause high values across the signal, even though the majority of high values are much smaller than the maximum decay time constant.

The opposite is true in a signal with one or two extremely large decay time constant

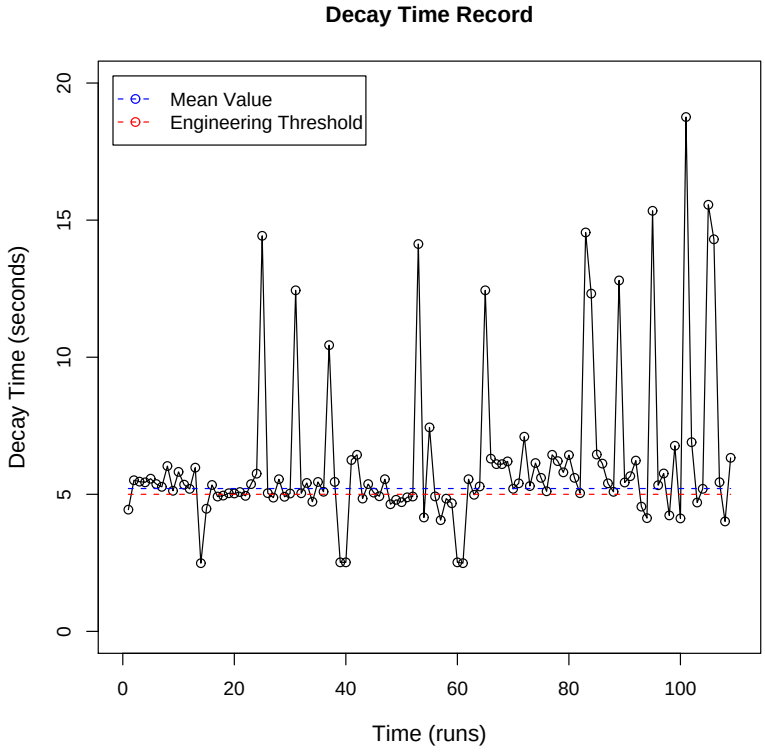


Figure 6.3: High Variance decay time constant record showing engineering threshold and decay time constant mean

values which will serve to drive the mean value up thus causing genuinely large decay time constant values to be assigned 0. In the response variable case, the goal of thresh-

olding is to ensure that decay time constants that are poor in relation to the signal under consideration, are assigned a high value in the logic model. The same is true for the predictor values which are used in the logic regression.

Predictor values pose a more difficult problem in terms of thresholding. Again, the mean value of the output can be used as the threshold but the same problems exist with the signal variance and the length of time the signal spends at maximum output. In addition, every generator is different and there is no general consensus on generator thresholding in terms of an individual generator's effect on damping.
A generator at half load may have an effect on a mode whilst another generator may only affect the same mode when it is operating at full load. The difficulty here is that by setting a threshold at half load or less across the generators, those machines that only affect damping at high output will reduce the model accuracy as they will have a high proportion of 1's for values at approximately half load or less. This is shown

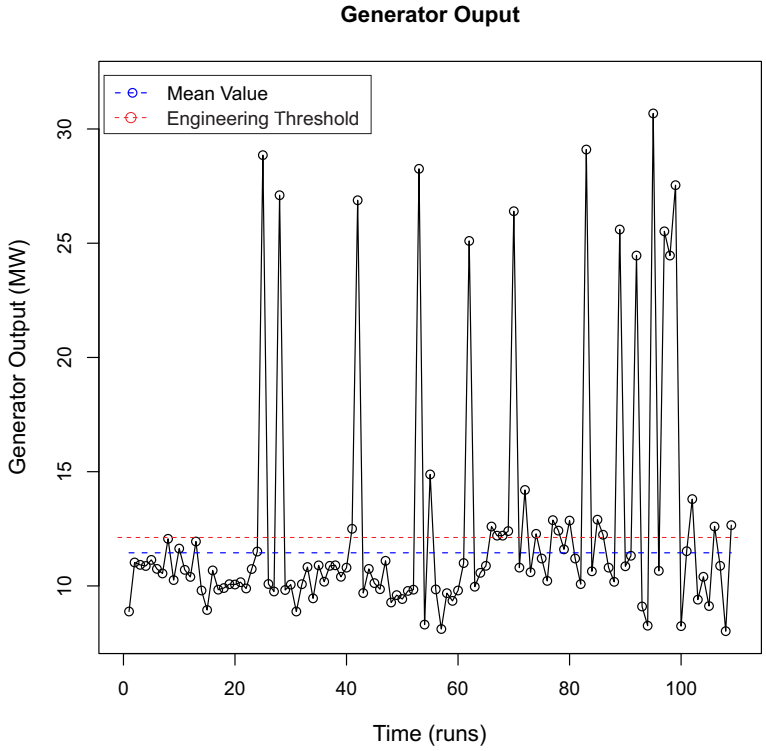


Figure 6.4: High Variance Generator MW record showing engineering threshold and decay time constant mean

in figure 6.4 and is similar to the decay time constant thresholding plot in figure 6.3.

Here, it can be seen that even the relatively low generator outputs are assigned high binary values as they are above the mean. In this case, the model accuracy can be reduced due to misclassification caused by these binary values.

In both the response and the predictor cases, thresholding values have to be assigned that ensure the relatively large values of decay time constant and active/reactive power are given a binary 1. Ideally, the threshold value should depend on the magnitude of the dynamic changes in the signals as well as the length of time the signal spends above a pre-determined engineering threshold. This ensures that a reasonable number of high values will be present in the model so that the interesting interactions can be determined. However, too many high binary values can create misclassification problems in the logic model meaning the fit will not be as good as the well thresholded model.

The wavelet coefficient thresholding procedure as well as the raw predictor thresholding is slightly different from the raw response variable case where the are no negative values to deal with. In this regard, both the raw predictor variables and both the wavelet transformed response and predictor variables are treated in the same way as follows.

Each data column is tested for negative values and subjected to the variance and absolute threshold tests explained in section 5.6.3. In the case of wavelet transformed variables there will always be some negative values if the signal is dynamic. There are usually some negative values in the raw data which depends on the conventions used for power flow and the operating point of the generators.
If the signals pass the absolute mean and variance tests and if negative values are detected then the data vectors are split into positive and negative values and the means of each positive and negative set calculated. By engineering judgment, the predictor thresholds were set at 1.4 times the predictor means as this seemed to work well through trial and error. This value was used for both the wavelet transformed and raw predictors.
The decay time constant thresholds are more important and since there is only one response variable, it was possible to apply more fine tuning to this threshold value.

However, the trial and error process always began at 1.4 times the mean with a final value usually between 1.2-1.5 times the mean value.

It must be noted that an adaptive thresholding technique is critical for the implementation of the source location methodology. The logic regression accuracy depends on the correct assignment of binary 1's in the transformations. An adaptive thresholding technique could be used to automatically fine tune the predictor and response variables to achieve the optimum model score. Although this is currently out of the scope of this PhD, it is a problem that will require a solution before implementation. Figure

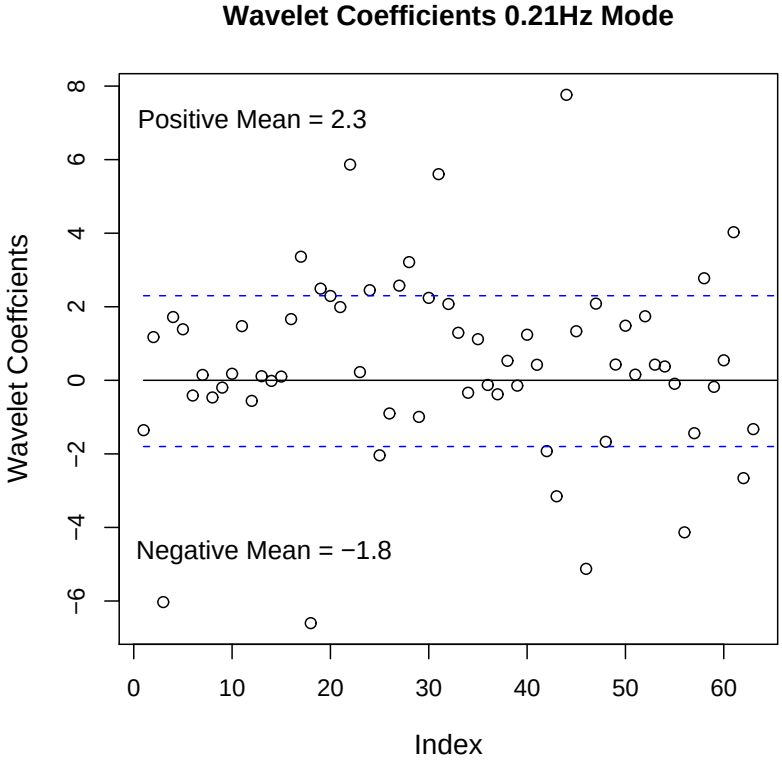


Figure 6.5: Wavelet Coefficient plot of a section of 0.21Hz decay time constants showing both mean values

6.5 shows a plot of the wavelet coefficients from a section of 0.21Hz mode decay time constant. Means for both positive and negative values are required as an absolute mean would be close to 0 due to the equality of positive and negative magnitudes in the wavelet transform.

6.3.2 Procedure for Logic Regression Analysis

The following procedure was used to determine the interactions in the power system that make poor damping more likely. A large section of this was covered in chapter 5 in section 5.5.2 but has been extended to include the logic regression procedure as well.

- Select a mode from the system and express damping in both Mode decay time constant (τ) and Damping Ratio (ζ)
- Select a suitable timeseries during which an event occurs and make sure there is sufficient shift in the data to account for delay in dynamics output (real-time data only)
- Build a matrix of independent (predictor) variables comprised of data series of active and reactive powers
- Filter data and select valid observations. Negative or undamped cases may be included if ζ is used as a damping expression
- Perform variable selection procedures on the data to determine useful predictors
- Develop a wavelet transformed GLM with either mode decay time constant or Damping Ratio depending on which is better
- Determine the variables that significantly influence the prediction by calculating the deviance statistic
- Develop logic regression models using both raw and wavelet transformed values. Use the deviance statistic respectively to score these models
- After the cross-validation steps etc. determine the best model in both the raw and wavelet transformed cases
- Score each predictor reported as significant in the logic models by tree reduction, WGLM or mutual information comparison
- Test interaction effects to determine if they are linked to mode damping

This procedure is standard for every logic regression model being developed. Depending on the number of trees in each model and the outcomes of the cross-validation steps etc. it may be that a number of trees and models need to be tested to determine the best interactions that can be used to control damping.
In order to fully test the methodology, the individual predictors that makeup the significant interactions have to be tested against participation factors and the dynamic model to see if there is a good match. This will also be done for the more casual predictors that are inevitably reported by the logic step.

Tree reduction can be used to determine the most significant individual predictors and these will be compared to corresponding results from the WGLM and mutual information models. The tree reduction step works by determining the largest tree(s) from the model that have large enough logic regression coefficients so that they are considered important. The size of the trees depends on the number of predictors in the model and can be judged on the number of generators that are likely to affect damping.

By reducing the number of leaves generated for each tree and then iterating the model, the least significant variable can be determined by comparing the trees before and after each iteration, as the least significant variable will be removed. This can be repeated until there is only one variable or "leaf" left in the tree.
The tabulated results from the tree reduction can then be compared to the WGLM and mutual information to see if there is an agreement between the different methods. In any case, these methods can be aggregated to establish a list of predictors in terms of their significance.

6.4 Preliminary Results from Logic Regression

Chapter 5 discussed the use of GLM's to model various response variables with a range of distributions including Poisson, binomial and gamma distributions. It was shown that by using the Gaussian distribution (general linear model), a well fitted model could be produced with the wavelet transformed data and significant predictors from

the GLM could be used to control the mode damping.

In the case of logic regression, if the response variable is a dichotomous binary value whose distribution is based on the *probability* of an event happening i.e. the probability of getting a 1 or poorly damped event, then logistic log-likelihood can be used to fit the model.

The parameter of interest then becomes the odds of belonging to class 0 versus class 1. This type of model was covered in section 5.4.3 and is based on binomial regression. In the logic case, the response variable is modelled as the *log of the odds* which is known as the logit of the probability or $logit(\mathrm{p})$. This is the canonical function of the binomial generalized linear model.

If the response variable is continuous (with binary predictors) then we attempt to model the means of the responses in two different subpopulations. This ties in with the classification problem mentioned earlier in that the model attempts to establish a Boolean statement that describes both classes. This is achieved by calculating the fitted values of the β coefficients when the Boolean statement is first true and then false.

It was shown in chapter 5 how the wavelet transformed decay time constant distribution resembles that of the normal distribution. In the case of the Icelandic 0.21Hz mode it appeared to be normally distributed with parameter values of μ=0 and σ^2=2. This enabled the use of a GLM to model the network as the wavelet transformed data was more linear than the raw data.

In this chapter, the wavelet transformed variables will be used in a logic regression model with a iteratively reweighted least squares fitting algorithm. This is due to the enhanced linearity of the wavelet response and its near normal response variable distribution. As a result, continuous wavelet transformed response variables are used as opposed to binary values with the model coefficients determining which interactions are significant in predicting both subpopulation means.

Also in chapter 5 it was discussed how a large proportion of the raw decay time constant values have no discernible distribution, hence mutual information was used to find and

remove negligible predictors. Non-parametric regression was also mentioned as a way to develop a statistical model without any assumptions about the response distribution. However, in this chapter the raw variables will be transformed into binary values with a binomial distribution for use in a logistic logic model. The motivation behind the binomial model stems from the notion of the logistic relationship between the response variable and the generator predictor variables in the power system. Figure 6.6 shows a comparison between the logistic function and typical τ vs. generator active power relationship for a power system model. The similarities between the two plots coupled with the binomially distributed response data means that the binomial raw data is suitable for logistic regression. As a result, the raw data is used in a logic model which is fitted using the logistic log-likelihood with the regression coefficients determining the interactions that make poor damping more likely.

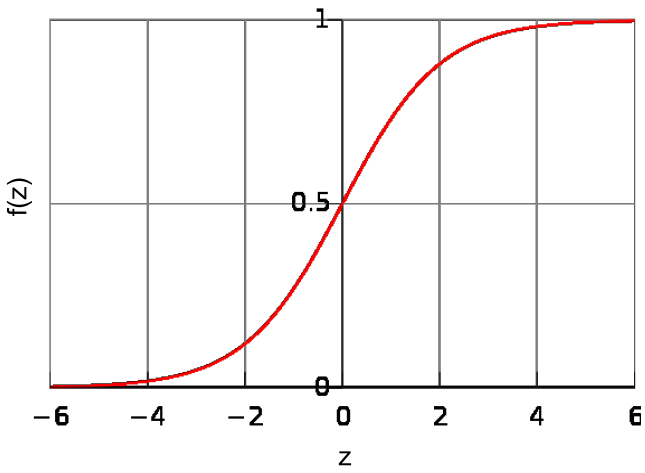
6.4.1 Development of Wavelet Transformed Logic Model

The development of the wavelet transformed logic model takes place in two parts. First, the logic model is developed with the full wavelet frequency spectrum in order to determine significant interactions in the system. The prediction capability of the full spectrum interactions is then be compared with that of the optimal scale logic model which uses only the interesting frequency components of the signals.

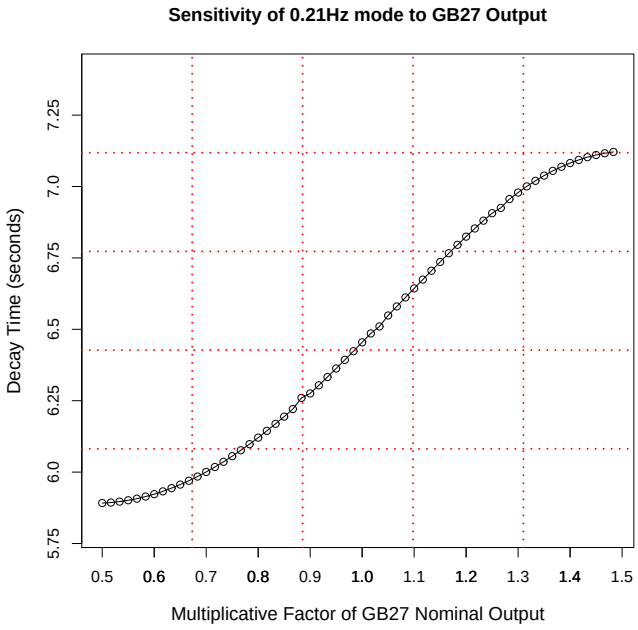
The variable selection methods outlined in chapter 5 will be employed throughout this chapter so as to improve the logic model prediction. Once both wavelet transformed logic models have been run, sensitivity analysis will determine if the significant interactions can be used to control damping. It is hoped the optimal scale models will outperform the full models as some of the redundancy will have been removed.

6.4.2 Development of Logic Expression for Icelandic Model

In this subsection, the 0.21Hz Icelandic mode was investigated using logic regression to determine a more robust analysis of the poor damping of this mode. The 256 length wavelet transformed dataset was used in this study as this setup produced the best results in the wavelet transformed GLM from section 5.9. As previously mentioned,



(a) The logistic function



(b) Sensitivity of 0.21Hz mode to GB27 active power output

Figure 6.6: Comparison of logistic function and response predictor relationship from the power system

the wavelet transformed data lends itself to a linear logic model and as a result the raw wavelet transformed response data is used with binary predictor data.

Initially, the active power variables were used to develop the model before reactive power variables are incorporated later on. Variable selection has already been applied to the model in chapter 5 section 5.11.3 with the remaining variables being used in this study.

Table 6.1 shows a summary of two of the interesting interarea modes from the Icelandic

State Variables					
Mode No	Frequency	τ	PF	LR RAW	LR WT
318	0.26Hz	12.1	e'_{q13}, e'_{q12}	e'_{q12}, e'_{q13}	e'_{q14}, e'_{q12}
322	0.21Hz	22.2	e'_{q11}, e'_{q24}	e''_{q30}, e'_{q11}	e'_{q11}, e'_{q24}

Table 6.1: Comparisons of model scores for different combinations of 256 length resolution levels for 0.21Hz Iceland mode

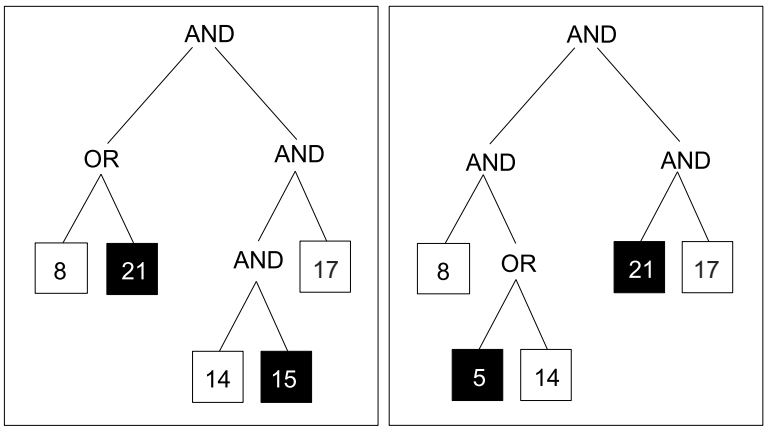
model. Information regarding mode frequency, decay time constant (τ), participation factors (PF) plus raw and wavelet transformed (WT) variables reported as the most significant in the logic regression (LR) are presented. This table can be compared to table 5.5 where the participation factors were in partial agreement with the statistical β coefficients. In table 6.1 the significant logic regression coefficients have been determined after variable selection but transformed back to a pre-variable selection index for comparison to the participation factors.

It can be seen that there is quite good agreement between the participation factors and the logic regression but that there are also some discrepancies that have to be explained. In chapter 5 section 5.8.1, it was shown how a wavelet transformed GLM was used to highlight significant variables that occasionally showed a higher correlation with damping than those determined via participation factors. This will be investigated later on in this chapter when more detailed results are extracted from the logic regression for the 0.21Hz interarea mode.

Figure 6.7 shows a comparison of logic trees derived from the initial runs for the 0.21Hz

data. Subfigures 6.7(a) and 6.7(b) show two of the logic trees produced from the same data. Since logic regression relies on a number of algorithms to search through the state space there is always some variability in the logic functions that are output from the study. This is due to small differences in the selection and transition probabilities in the models that lead to slightly different logic trees being produced.

This variability is very important as it allows all the possible important states to be collected and tested for significance. This is an unavoidable attribute of logic regression though there are ways of determining the significant predictors and interactions from the various models. These techniques are covered later in this chapter. The wavelet transformed logic regression models in figure 6.7 are in good agreement



(a) Five leaf logic tree output (b) Five leaf logic tree output

Figure 6.7: Comparison of outputs from wavelet transformed logic regression step

with the wavelet transformed models in chapter 5 in section 5.9. The generator variables reported in subfigure 6.7(a) correspond to generators G8(GB27), G14(GB52), G17(GB102), G19(GB117) and G21*(GB138*) (G21* is a *not* generator which means G21 is at low output, as defined by the threshold set to classify the data into a binary 1 or 0, when damping is poor). So G8 is generator number eight connected to bus 27. The numbers in brackets correspond to the bus numbers the generators are connected to in the Icelandic system model. From here on in, generator number will be used to define particular generators highlighted in the statistical models. Generator numbers

can be referred to the bus names as shown in the single line diagram of figure 4.2. However, bus numbers such as GB27, which corresponds to the bus G8 is connected to, will occasionally be reported with their generator number in the rest of this work.

For example, we see that generators 8,14,17,19 and 21* are highlighted by the logic regression model. These generator numbers correspond to generators in the model connected to buses at Blanda (BLA), Krafla (KRA), Sigalda(SIG), Burfell (BUR) and Iranyjafoss (IRA). Generators 5 and 15 correspond to buses at Mojvidenni (MJO) and Marmajno (MNO). In reference to figure 4.2, the location and thus generation patterns can be seen in the single line schematic.

A *not* generator refers to a generator that outputs below its selected mean during periods when the mode of interest is well damped. Essentially, if a mode is poorly damped then the related generation patterns (Boolean statements) are determined by the logic regression in terms of binary numbers (0 or 1). If a generator is consistently below its engineering mean i.e. is 0, then it will be included in a particular pattern but will be assigned as a *not* or "not on" generator.

The generator variables reported in subfigure 6.12(b) correspond to generators connected to buses 27(G8), 52(G14), 99(G16), 14(G5)* and 138(G21)*. It can be seen that there is agreement between both the wavelet model as well as the GLM in chapter 5, although there is some variation in the raw and WT logic models. However, both models are required as they can produce mutually exclusive variables which can be used to control damping. The next step is to develop a method to extract these significant interactions and predictors from both classes of logic model.

6.4.3 Reduction and Validation Methods

As discussed previously, there is some inherent variation in the logic regression models caused by variations in selection and transition probabilities. In order to establish the correct variables to keep in the models some method(s) have to be developed to ensure all of the important information describing the model is retained.

In the previous section we discussed the use of the WGLM, mutual information and tree reduction as a means to determine the most significant variables. As well as expanding on these methods we will also look at how cross-validation and Markov Chain Monte Carlo (MCMC) algorithms may be used to determine the significant variables.

Mutual Information

In section 5.12.3 the mutual information for the Icelandic 0.21Hz mode was plotted in figure 5.35. It can be seen in this plot that of the reported variables 11,21,24,25,27, which correspond to generator variables 8,14,16,17,19 (after variable selection), all of them have significant mutual information coefficients with variable 25 (G17 or GB102) posting the lowest coefficient of the selected variables. However, it must be remembered that the mutual information step took place with raw coefficients and not wavelet coefficients, therefore some of the WGLM variables may not be reported in the M.I step.

Wavelet Transformed GLM

In section 5.12.3, the WGLM reported generators 8,16 and 17 (which are connected to buses 27,99,102) as significant variables that were linked to the mode damping. These have also been reported in the logic regression trees though they haven't all been reported in the same simulation. This is why a validation method has to be developed as it is quite common for significant variables to be spread across a number of simulations.

Tree Reduction

Figure 6.8 shows one possible evolution of a logic model when one leaf is removed each time the logic regression is run. In this case, the five tree model is used as a starting point. This is shown in subfigure 6.8(a). With one leaf removed the model is run again, with the output shown in subfigure 6.8(b). Here the logic tree has retained most of its shape and all the original variables except generator G14 which is connected to bus 52, which corresponds to the Krafla bus. In addition, variables G18 and G21 are combined in an OR statement and is AND'd with G17. However, this still retains the

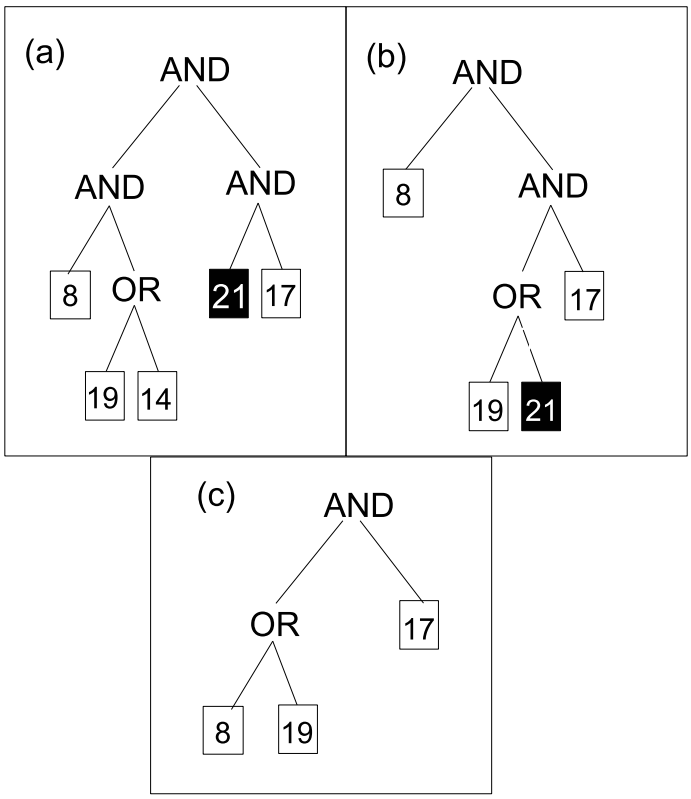


Figure 6.8: Tree reduction from wavelet transformed logic model

most significant variables as variables G8,G17 and G19 or G21* were reported in the four leaf regression.

Subfigure 6.8(c) shows the final logic tree with the number of leaves reduced to three. Here, the shape of the tree is quite similar to subfigure 6.8(b) except that the two significant variables G8 and G19 have now been OR'd together which means that only two variables will be reported in the final result. This change of operator in the tree reduction is not uncommon and is caused, again, by small differences in the state space paths, facilitated by the change of constraint i.e. reduced tree size.
It is clear that tree reduction alone isn't suitable for establishing significant variables in the model as there is some variability in the models themselves. It does however, give us an idea of which variables are most significant as these tend to be retained, although the actual interactions are not as clear as in some cases the Boolean operators vary.

Additionally, mutual information and the WGLM tell us about the significance of each variable but we rely on the logic regression itself to determine the Boolean expression that governs the model. However, these steps are useful as they can indicate which variable to change first in order to control damping. Since the model is subject to variability, some method that involves an aggregate or average set of variables must be derived to determine the combinations of variables that affect the damping. Once this aggregate expression is derived, each component can be tested for significance in relation to mode damping.

6.4.4 Model Search Methods

Since the number of possible Logic Models we can construct for a given set of predictors is huge, we have to rely on some search algorithms to help us find the best scoring models. In addition to the search algorithms, different methods may also be used to find optimum model sizes and runs that can be used to determine the lowest model score. The following section details some of the methods used to achieve this.

Cross-Validation

The primary goal of cross-validation is to determine the optimum model size for the data available. This is required as different model sizes produce different fits for the data. In the logic regression case, the best scoring model generally overfits the data. Cross-validation is used to access this over-fitting of large models.

Cross-validation involves partitioning the data into complementary subsets, performing the logic regression analysis on on one subset (*training set*) and validating the analysis on another subset of data (*testing set*). To reduce variability, multiple rounds of cross validation are performed using different partitions and tree sizes, and the validation results are averaged over the rounds. The cross-validation step is fairly comprehensive in that it runs a large number of models with various data lengths. By doing this, the optimum model size can be found while taking into account various data lengths which can effect model performance. This cross-validations step thus results in a better model prediction. Figure 6.9 shows a plot of the model scores for the cross-validation

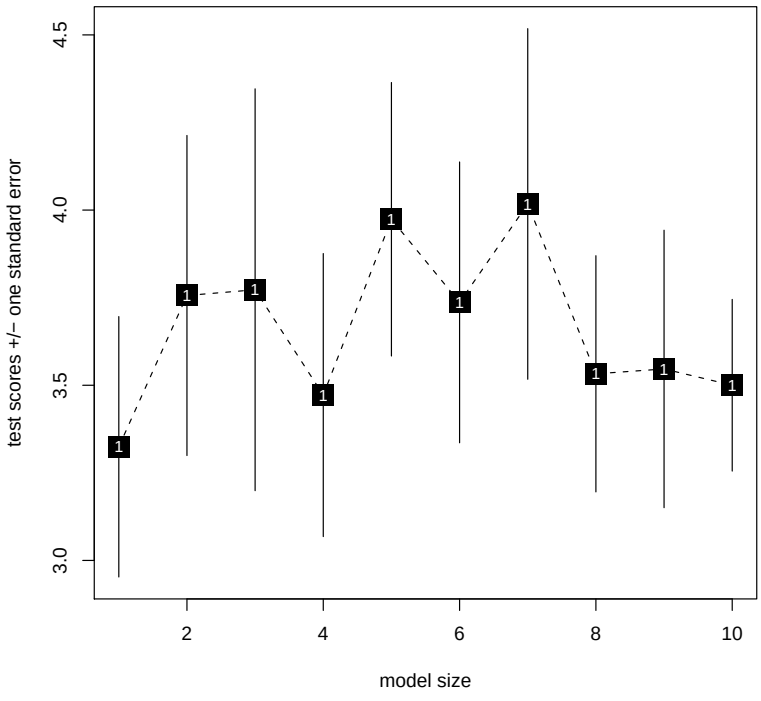


Figure 6.9: Cross-Validation step showing the model scores for the number of leaves (variables) in the model

step performed on the 0.21Hz Icelandic data. It can be seen that a model size of 4 produces the lowest scoring model as a model score of 1 is null and void as it is a single interaction. This means that every model from here on should be built with four leaves with the goal of finding the combination of four variables with the lowest score.

Conditional Permutation Test

In searching for the best scoring model of a certain class the search will always find the best model but the question of whether or not the model fits the signal (our dependent variable) or noise (overfitting) still has to be answered.
For any model class considered in our methodology (linear regression, logistic regression etc.) we first find the best scoring model given the dataset. The null hypothesis that is tested states, "there is no signal in the data." If that hypothesis is true, then the best model fit with the response variable randomly permutated should yield the same score as the best model fit on the original data. However, if the model score is improved then the model fits the true signal and not the noise. This is due to the fact that a model fitting noise will return the same model score no matter how the response data is randomized or rearranged.

Figure 6.10 shows the histogram of best scores across all the possible model permutations up to a one tree, seven leaf model. It can be seen how the permutation scores improve until a model is conditioned with one tree and four leaves after which the scores do not change very much. This is further evidence, in conjunction with the cross-validation, that a single tree four leaf model is the best model for this data. It also shows that the model isn't overfitted. It must be noted that the conditional permutation models are scored slightly differently to standard logic regression scores, hence their difference in value. Information on the cross validation scoring procedure can be found in appendix B.

Greedy Algorithm

Figure 6.11 shows the scores for the one tree seven leaf logic model using the greedy algorithm. This was run in order to compare the simulated annealing algorithm per-

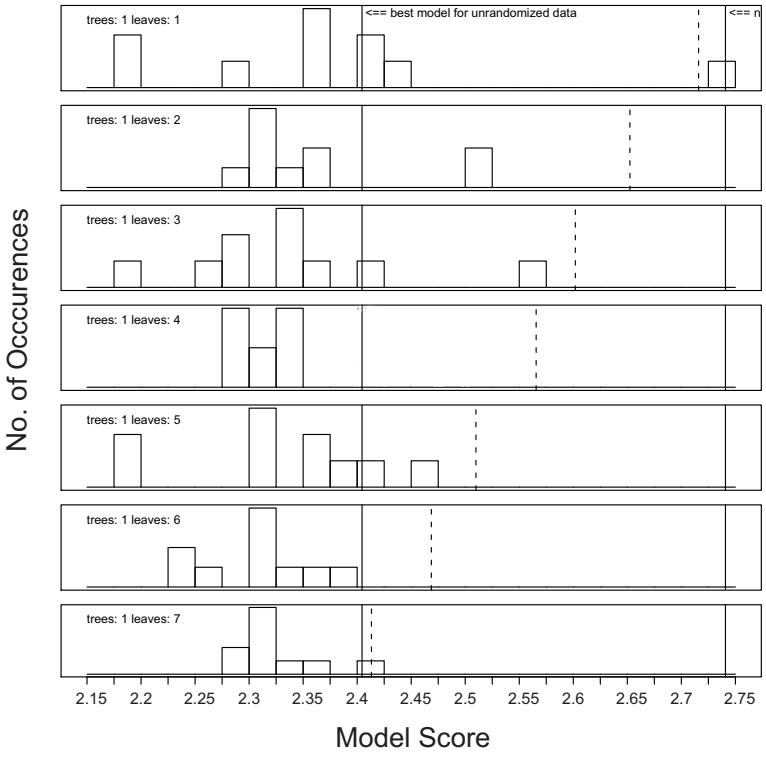


Figure 6.10: Conditional Permutation Plot showing how the scores improve until one tree and four leaves is reached after which there isn't much improvement

formance against that of the greedy algorithm. It can be seen in the figure that the greedy algorithm also terminates at a leaf number of four as no further improvement in the model is seen afterwards. Again, this enhances the case for using a model with

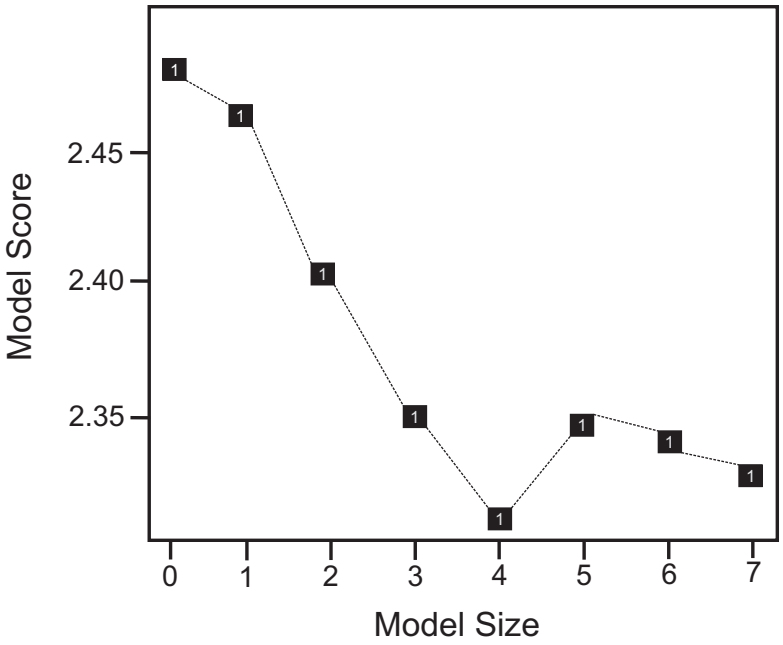


Figure 6.11: Score plot from the greedy algorithm with seven leaves.

four output variables of interest.

Monte Carlo Logic Regression

Markov Chain Monte Carlo (MCMC) methods which include random walk Monte Carlo methods, are a class of algorithms for sampling probability distributions based on the construction of a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a large number of steps is then used as a sample from the desired distribution. The quality of the sample improves as a function of the number of steps. This is what is done in Monte Carlo Logic Regression.

The algorithm used is a reversible jump MCMC algorithm which is described in the literature [80]. Other than the length of the Markov chain, the only parameter that needs to be set is a parameter for the geometric prior on the model.

Within the MCMC framework there are a number of outputs that allow each of the

model predictors to be assessed in terms of the number of their appearances in each of the MCMC models. These are summarized below:

Single: A vector with as many elements as there are binary predictors. Single[i] shows how often predictor i appears in any of the MCMC models. Note that when a predictor is twice in the same model, it is only counted once, thus, sum(size[,1]*size[,2]) will typically be slightly larger than sum(*single*).

Double: A square matrix with the size equal to the number of binary predictors. Double[i,j] shows how often predictors i and j are in the same tree of the same MCMC model if $i > j$. If $i \neq j$, double[i,j] equals zero. Note that for models with several logic trees, two predictors can both be in the same model but not in the same tree.

Triple: Square 3D array with the size equal to the number of binary predictors. See double, but here triple[i,j,k] shows how often three predictors are jointly in one logic tree.

Thus, MCMC regression can prove very useful, as significant variables can be tracked across a large number of different models to determine which other variables they often appear in conjunction with.

6.4.5 Results from Logic Regression

In this section, results are presented from the logic regression based on the four leaf models derived from the cross-validation steps etc. The results are then tested against an eigenvalue analysis as well as against comparisons of the normalized decay time constants versus the normalized predictor values over a suitable time window.
The results are derived from the Icelandic 0.21Hz mode which has already been subjected to variable selection and pre-processing, which leaves a total of 23 generators that are used in this study.

This model is based on the optimal scale combination detailed in section 5.9 with a 256

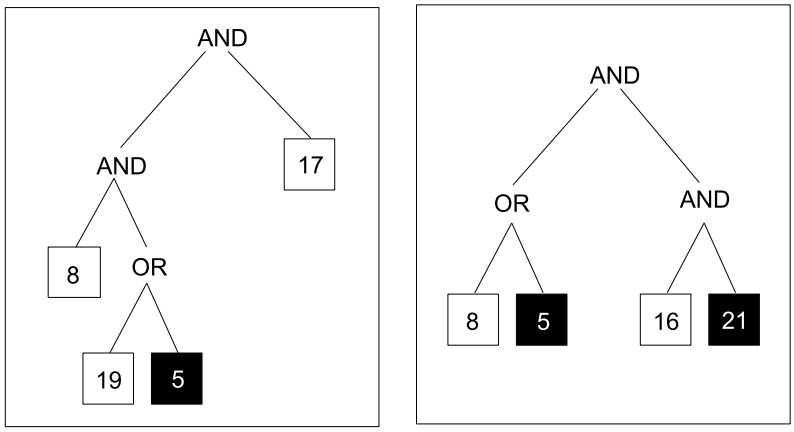
data length. This optimal scale GLM used resolution levels 3,4,6 and 7 as these levels had the most significant wavelet coefficients when the signal was deconstructed. This optimal scale combination was shown to produce the best fitting GLM's in chapter 5 and as a result, were used in this study to determine the best logic models. Table 6.2

MCMC Run		
Model	Levels	Score
1	1,2,3,4,5,6,7	3.524
2	1,2,3,4,6,7	2.513
3	4,5,6,7	3.036
4	3,4,6,7	2.301

Table 6.2: Comparisons of model scores for different combinations of 256 length resolution levels

shows the comparison of the best scoring models for the logic regression step using the various optimal scale combinations that were tested for the WGLM's. It can be seen that model 4 which consists of the resolution levels 3,4,6 and 7 performed best as these were the resolution levels that also produced the best fitting WGLM's. The scoring used for both the raw and wavelet transformed logic regression models in this chapter is the statistical deviance score which is based on the maximum likelihood estimators. In this case, as with the WGLM's, the lower the deviance score the better. This scoring method is covered in 5.4.3 and is derived from maximum likelihood estimators (MLE). This result makes sense as most of the interesting dynamics occur at these resolution scales. This means that the specified resolution levels of the response and predictors are quite well correlated, in comparison with the correlation of the redundant resolution levels.

Figure 6.12 shows a comparison of two four leaf models derived from the Icelandic 0.21Hz logic regression model. It can be seen that the outputs vary quite a bit, with a total number of six variables or generators being represented between the models. The MCMC step allows all the variables in each model to be tabulated and compared against each other to determine how many times they appear in the same model. This in turn allows us to establish the most common interactions produced within the model as well as removing some of the redundant variables and interactions that also take



(a) Four leaf logic tree output (b) Four leaf logic tree output

Figure 6.12: Comparison of four leaf output from the logistic regression showing some of the variation in the model with different generator combinations being outputted. See figure 4.2 which references generator numbers to bus locations in single line diagram

place.

Table 6.3 shows the comparison of the number of observations of joint variables in

MCMC Run		
Single	Double	Triple
8	21,16,17,14,23,22	16,17,14,20,1
16	8,14,21,17,21,23	14,21,17,19,22
21	8,16,1,14,17,5	16,17,14,20,1
14	16,21,8,23,19,18	8,21,6,17,3
17	21,8,14,16,8,7	8,14,16,1,19
19	14,16,8,21,10,3	16,8,23,21,6

Table 6.3: MCMC Comparisons from the most common generators derived from the models

the MCMC models. The single column shows the six of the most common variables in the MCMC step with generator G8 being the most common. The double column displays the most common variables that appear in the models alongside the single variable in column 1. For example, the double column in row 2 contains the generators 8,14,21,17,21,23 which means that generator 8 (corresponds to Blanda, see figure 4.2) is jointly observed with generator G16 (corresponds to Rangnar, see figure 4.2) across the monte carlo simulation more than any other variable. Generator 14 is the next

most common variable jointly observed with G16 an so on.

In addition, by taking a sample of 10 or so models it is possible to determine the most common operator between variables. For instance, if we take 10 models where variable 8 and 21 appear, we discover that 100% of the time they are connected via an OR operator which means they each have their own separate effect on damping and do not combine to exert any influence on the mode.
The triple column contains the most frequently observed variables in the MCMC models that also contain the single variable and the most common double variable. For example, row 1 of the triple column contains the generator variables 16,17,14,20,1. This means that these variables appear more commonly in models containing generator variables 8 and 21 than any other variables with generator variable 16 being the most common partner of generator variables 8 and 21.

The MCMC single variables seem to be in good agreement with both the logic regression results shown in figure 6.12 as well as the results from the wavelet transformed GLM (section 5.9, table 5.5) with G8(GB27) and G16(GB99) being reported as significant. A large number of the common single variables also appear jointly in the model with other common single variables as evidenced in the double column.
For example, generator variable G8 also appears quite regularly with generator variables G21 and G14 which are also frequent single variables. In contrast, it can be seen that generator variable G8 is also jointly observed with variable G23 even though variable G23 is not a high ranked common variable. G23 corresponds to Steingmar (STE) in figure 4.2.

This is the benefit of the MCMC step which can be used to highlight these coincidences across a number of models. This allows the less frequent variables to be included in the model by virtue of association with more common variables.
This is illustrated with variable G17 as it hasn't been reported as an important single variable (ranked fifth) but is present in quite a few models containing significant variables. As we will see later, these less common variables can have a significant influence

on damping when they are present in certain interactions.

The triple values also offer some further information on the frequency of groups of variables in the model. In the case of the significant generator variables 8 and 16, it can be seen that generators 17 and 22 are obsevered fairly often with this pair. In this case, it may be useful to look at the effect these combinations of generators have on damping. By using these double and triple values, results can be obtained across all the possible models which means that no information is lost in the logic model.

The use of this aggregate method spanning a large number of models is ideal for the source location problem due to the large number of interactions that take place in the power system. Although, some variables are more common than others, it is their interactive effects that matter and as a result no significant variable produced from the MCMC step (single, double or triple) can be discounted as they may prove to be very significant in terms of their interactions.

6.4.6 Reliability of the Icelandic Logic Model

The reliability of the Icelandic model in this section will be tested and validated in a number of ways. As in chapter 5, participation factors derived from eigenvalue analysis will be used to compare the logic regression results with the dynamic performance of the system to see if they can be reconciled.

In addition, plots of the significant variables from the logic model will be compared with the mode decay time constant to determine if there is any correlation present. Since the logic model focuses on the interaction effects, the dynamic model will be modulated to determine the extent to which, these interaction effects act on the mode damping. Figure 6.13 shows a typical logic regression tree from the Icelandic model. The logic models are subject to some variation and as a result, the reported variables may differ from simulation to simulation. In the previous section, methods were discussed to overcome this difficulty with the use of Monte Carlo methods to determine common groups of variables.

It must be noted that the MCMC step is used to supplement the logic models. Figure

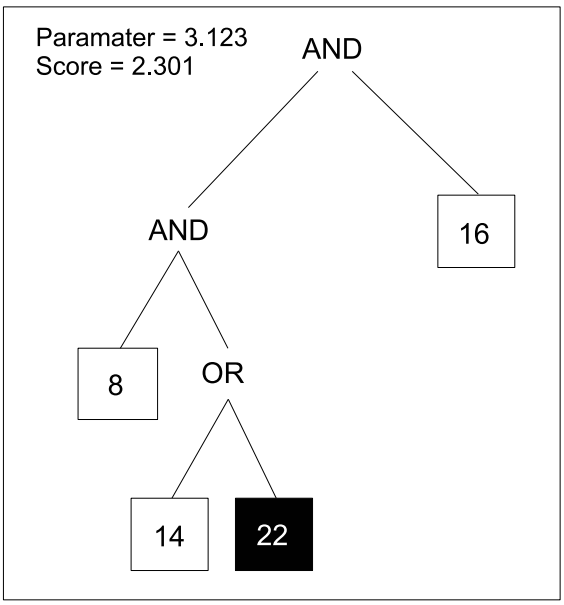


Figure 6.13: Optimal Logic Tree from Icelandic Model with Score and Parameter Values

6.13 displays the score and parameter value for the tree in question. The parameter value refers to the logic regression coefficient assigned to each interaction in the model. In this case only one tree has been specified which means there is only one parameter value. The score function is discussed in appendix B and defines how well the logic model has been fitted.

The wavelet transformed Icelandic data lends itself to a well fitted logic model. As a result the variation in the outputs from run to run is minimal. However, some variation does exist and as a result, aggregate methods are still required. Once again, generators 8 and 16 appear in the logic tree in figure 6.13.
These variables were shown to be significant via the MCMC step as they appear in an appreciable number of Monte Carlo models. In addition, generator variables 14 and 22 also appear in the tree. G22 refers to A12 in figure 4.2. Although G14 was shown to be common in the MCMC model, G22 has been included as it appears regularly in the models whilst in the presence of certain other variables. Therefore, these variables are a good starting point in determining which predictors and interactions are strongly linked to mode damping.

Table 6.4 exhibits the MCMC results for the variables reported in the starting logic

MCMC Run		
Single	Double	Triple
8	21,16,17,14,23,22	16,17,14,22,5
16	8,14,21,17,21,23	14,21,17,19,22
14	16,21,8,23,19,18	8,21,6,17,3
22	16,8,23,14,15,5	8,14,23,5,21

Table 6.4: MCMC Comparisons from the variables derived from the initial logic tree in figure 6.13

regression. The point of this section is to compare the logic trees and their affiliated variables with the dynamic participation factors to see if they can be matched. Additionally, the interactions will be tested in the dynamic model by resolving their effect on mode damping when some of the participating generators are adjusted. Figure 6.14

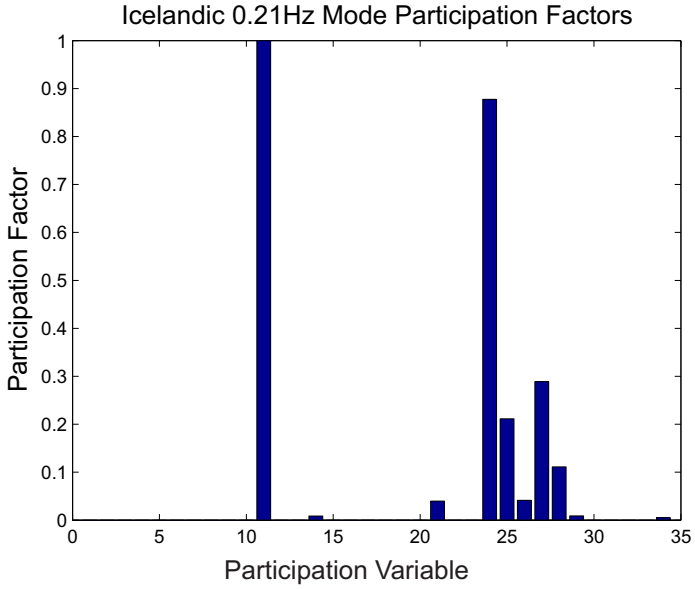


Figure 6.14: Icelandic 0.21Hz Mode Participation Factors derived from rotor angle states. Participation variables refer to the original 34 generators in the model, whereas in the text, each participation variable is referenced to a generator number as the number of generators where reduced as a result of variable selection

displays the average participation factors for the Icelandic 0.21Hz mode under consideration. The results have been normalized with the largest value attributed to variable 11 or G8. It must be noted that the 34 variables in the plot represent the 34 generators

in the model as the slack generator has been removed from the statistical model. However, the variable selection process removed additional generators from the model and as a result the indexing of the generators used in the logic regression and wavelet transformed general linear models has changed. However, when referencing figure 6.14, generator numbers are given as well as the participation variable number e.g. Participation variable 11 is equal to generator G8 in the Icelandic power system model.

As expected, there is some agreement between the participation factors and the logic regression step. These results are similar to those in chapter 5 which compared the participation factors to the WGLM results and found reasonable agreement. In this case, variables 11 and 24 (G8 and G16) are clearly the highest participating machines in the system. Other machines that have a higher than average participation include G17 and G19 as well as G21.

However, some of the commonly reported generators from the MCMC step have negligible participation factors relative to the highly participating machines. The focus of this section is to determine the effect these machines have on the damping in conjunction with the more obvious generators from the WGLM. This must be studied as there is some disrepency between the logic regression step and the participation factors, the source of which may lie in the interaction effects of the generators.

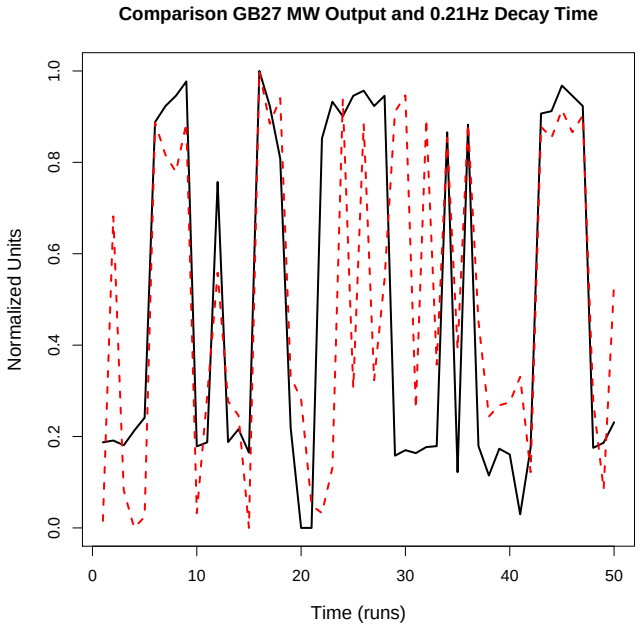
In the previous logic regression analysis we presented a number of interactions that seem to have an effect on the mode damping. Together with the cross-validation and MCMC steps it is possible to produce a list of variables in order of their importance. This rank is based on the number of times they appear in the model as well as the number of appearances in conjunction with significant (single column) variables. Table 6.5 shows the comparison between the aggregated logic regression variables and the participation factors for the 0.21Hz mode. The first row of results also include negative Boolean variables i.e. GB138(G21) is a NOT value, whereas row 2 contains the ranked logic variables with the NOT values removed as we are only interested in active positively correlated generators. It can be seen that both sets are generally in agreement

MCMC Run		
Simulation	LogicReg	PF's
+/- Gen. Values	G8,G16,G21,G14,G17,	G8,G16,G19,G17,G20
+/- Gen.Bus Values	27,99,138,52,102,	27,99,117,102,126
Gen. + Values	G8,G16,G14,G17,G19	G8,G16,G19,G17,G20
Gen.Bus + Values	27,99,52,102,117	27,99,117,102,126

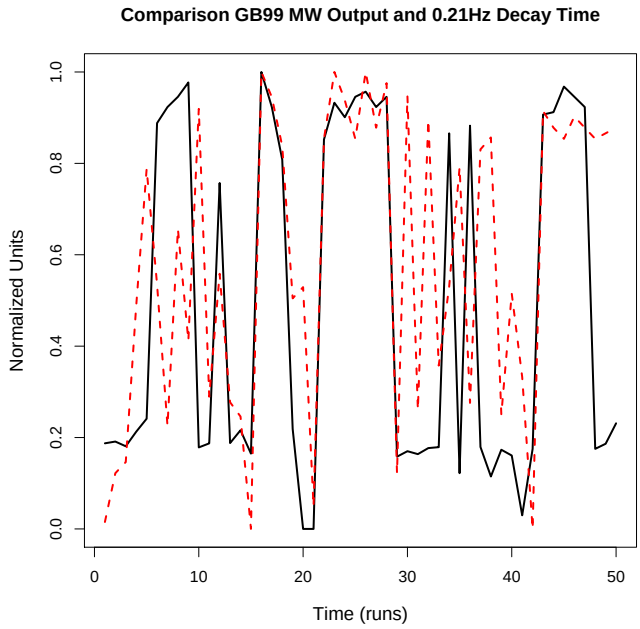
Table 6.5: Significant generator variable comparison between logic regression and participation factor results for 0.21Hz mode with both generator number and the generator bus they are connected to

except for the presence of generator variable 14 (GB52) in the logic analysis which has a relatively small participation factor (generator variable 21 in figure 6.14).
In reference to table 5.6 and figure 6.14, it can be seen that the WGLM of the 0.21Hz mode also failed to highlight this variable as it was deemed to have a negligible effect on the mode when considered as a one-way interaction.

Referring to figure 6.13 again which shows one of the most frequent interactions produced by the logic regression step, generator variable 14 (GB52) is part of a nested OR statement and should have an effect on damping. It would therefore be useful to determine the effect G14(GB52) has on mode damping as it is not deemed significant in the analytical model.
In figure 6.15 the subplots show the superposition of the 0.21Hz mode decay time constants against the generator outputs for G8(GB27) and G16(GB99). These were the two most significant machines reported in the logic regression and the WGLM step in chapter 5 section 5.8.3. In the window shown, the G8(GB27) and G16(GB99) outputs seem to correlate with all the events as either (or both) of them are high when the decay time constant is poor. By using a simple linear regression, single effects may be considered i.e. single predictor effects on damping. In the logic regression case, we can see that at least two variables have an effect on mode damping. However, the logic tree in figure 6.13 also suggests that G14 (GB52) has an effect on the mode damping. Subsequently, it may be worth looking at the effect a reduction in this generator output has on the mode damping. Figure 6.16 shows the decay time constant plots derived from the loadflow runs with G16 removed. G16 connected to generator bus 99 was



(a) Generator 8 (GB27) Comparison



(b) Generator 16 (GB99) Comparison

Figure 6.15: Plots showing significant generator outputs (red) against mode decay time constant for G8(GB27) and G16(GB99)

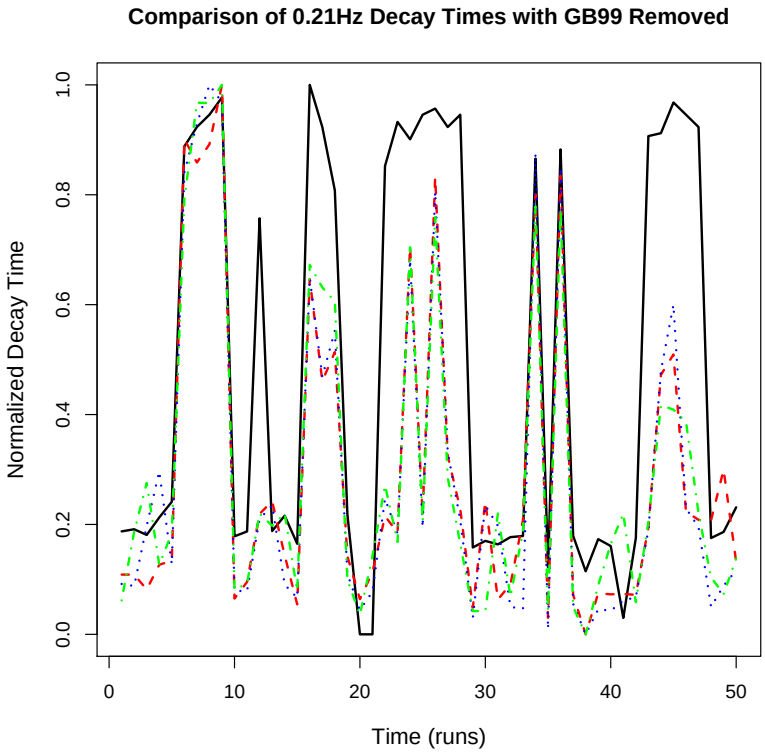


Figure 6.16: Comparison of selected decay time constant plots when G16(GB99) is modulated. The different colour plots refer to different loadflow scenarios run to produce different decay time constant plots

not removed completely as this would change the topology of the network. This could have a major effect on the system dynamics and render corresponding modes incomparable and untraceable. Instead of removing the generator, a constant power source was modelled according to the preset limits of the systems. For G16, the max/min outputs (limits) were set at approximately ±80% nominal value and as a result, the constant power source was set at 20% the nominal output of G16 in order to test the effectiveness of this generator.

The different colours on the plot refer to decay time constants derived from different loadflows of the Icelandic model. For the most part, the model produces quite stable and coherent decay time constant records.
The interesting point in this plot is the effect the modulation of G16(GB99) has on the mode damping. The reduction of G16(GB99) to a constant power source has caused a similar reduction in the decay time constant magnitudes across parts of the event. This is especially true in the regions where G16(GB99) is the main predictor such as the period between 20 and 30 seconds.

In section 4.6.2 we discussed the use of intentional variable adjustment i.e. the ramping of different combinations of generators at different times along the window. This has been applied in figure 6.16 as the reduction in G16(GB99) has only reduced decay time constants intermittently along the window.

However, the decay time constants have not dropped to the minimum value due to the action of G8(GB27). In the regions were both G8(GB27) and G16(GB99) are high in the original plot (figure 6.15) the decay time constant has roughly halved in value which shows that both G8(GB27) and G16(GB99) combine to cause the poor decay time constants observed and not just GB27.
The replacement of G16(GB99) has caused the dynamic behaviour of the mode to change indicated by differences in the decay time constant magnitudes. At various points in the window, the decay time constants have remained constant which suggests that other variables are responsible for these "mini-events".

In this case it could be either G8(GB27) or G14(GB52). This is testament to the complexity of the system as evidenced by the fact that the mode evolves over time and is thus altered by a number of predictors and combinations thereof as time progresses. However, it is clear that at least two variables alter this mode in combination, as opposed to a single predictor.

In comparison to G16(GB99), figure 6.17 shows the decay time constant comparisons when G8(GB27) has been replaced with a constant power source. The output was set at 23% nominal output in accordance with the system limits. Again, the decay time constant is seen to be reduced in the regions were G8(GB27) is the most significant predictor, although they have not been reduced to the minimum value of 3-4 seconds as seen when the system is optimally stable. It is evident that G8(GB27) is interacting with other variable(s) in order to perpetuate these long decay time constants.
If this were not the case, upon removal of G8(GB27) the decay time constants would drop to the minimum values observed in figure 6.17.

Figure 6.18 shows the selected decay time constant profiles when both G8(GB27) and G14(GB52) are replaced simultaneously with a low constant power source. This plot is the most revealing of all as the effects of simultaneously removing both G8(GB27) and G14(GB52) can be seen, with some of the events having completely disappeared from the profile. The first event from 4-10 seconds as well as the 34-36 seconds event are no longer present after the loadflow runs with G8(GB27) and G14(GB52) removed, even though they can still be partially observed when G8(GB27) is modulated alone (figure 6.17).

This suggests that the interaction of G8(GB27) and G14(GB52) has a strong effect on mode damping as suggested by the lowest scoring model tree in figure 6.13. Since these events are only diminished when both machines are reduced. This means that both these machines can be used to provide a more robust control of damping. Figure 6.18 shows that generator variables G8 and G14 are an AND pair, in that both of them (as opposed to either of them) have to be reduced in order to affect damping. This is

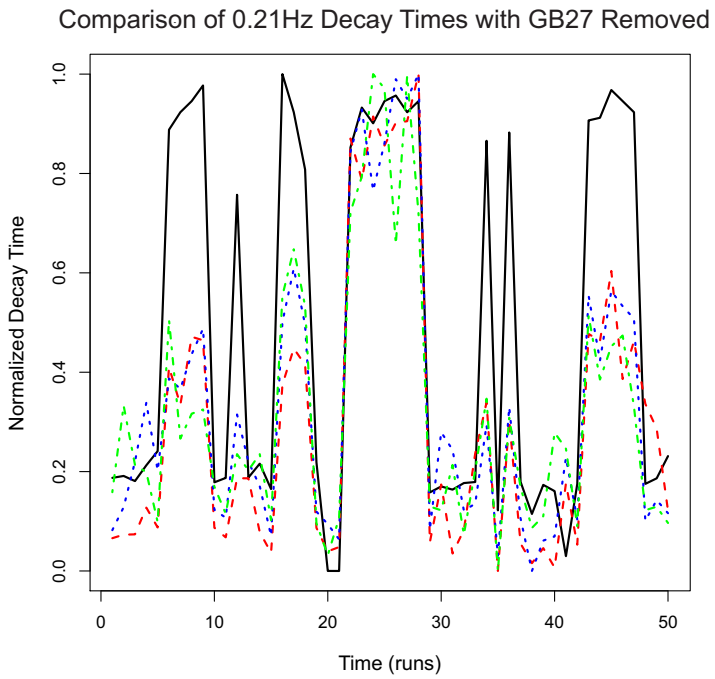


Figure 6.17: Comparison of selected decay time constant plots when G8(GB27) is modulated

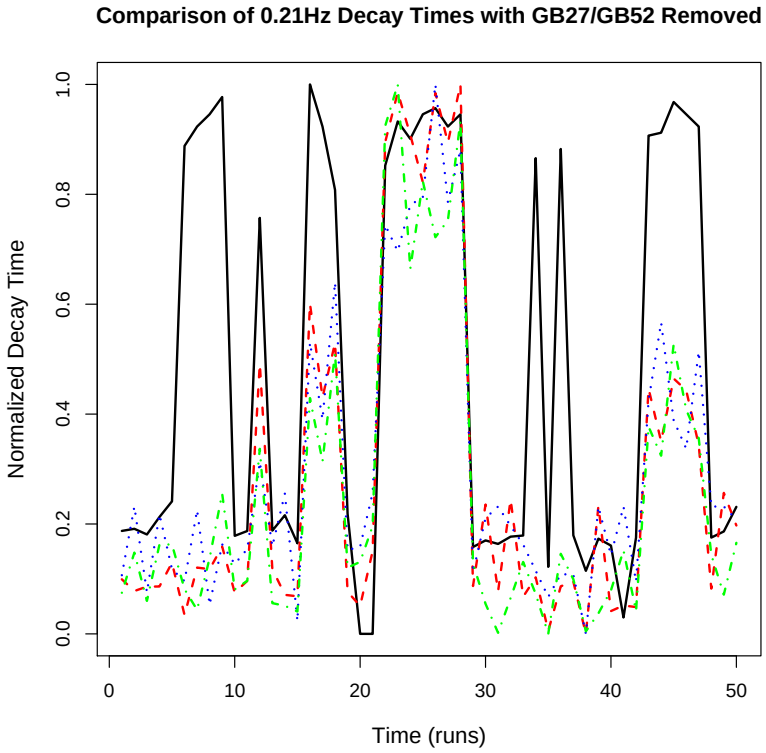


Figure 6.18: Comparison of selected decay time constant plots when G8(GB27) and G14(GB52) are modulated

further evidenced in figure 6.13 where variables G8 and G14 are AND'd while variable G14 is also OR'd with variable G22. From the sample model runs it is clear that when both of these variables appear in a model output i.e variables G8 and G14, the vast majority of the time they are in an AND pair.

It has been shown in this case that a number of variables are linked to damping across the window shown. 8(GB27) and G16(GB99) were shown to have the greatest effect as when either of the machines are off, decay time constants were partially degraded. When both machines are off the majority of the window is well damped and the number of "mini-events" reduces substantially. However, some of these events remain, as they are dependent on the interaction of G8(GB27) and G14(GB52) as opposed to any single machine. It is only by the reduction of both these machines can the long decay time constants be reduced. More complex modes would be governed by larger interactions which would expand the options available for an operator attempting to control the system modes.

The above example was provided as an exercise to aid understanding of the logic regression step. For a given length of data, the variables affecting the decay time constant at any instant, varies as we move along the decay time constant window. As a result, the logic regression step reports these variables in a logic tree.
Since the interaction effects evolve over time, the choice of data length is crucial so as not to overcrowd the model with a large number of interactions. The above example was perfect in that G8(GB27), G16(GB99) and their interactions controlled most of the window with some mini-events being more dependent on the G8/G14 combination. Both of these interactions were expressed in a compact logic tree without any cluttering which allows the interactions to be clearly understood.

In the MCMC regression we also observed other variables that had been proposed in interactions that supposedly affected mode damping. We have already seen two interactions thusfar that can be used to control damping but there may be a large number of other interactions that effect the mode as well. The logic regression step also re-

ported generator variables 17 and 19 (GB102 and GB104) as significant and there is a good chance that these variables also have an interaction effect on the 0.21Hz mode. However, it is clear that this methodology depends largely on the assimilation of an appreciable and complex set of interacting variables.

Figure 4.1 in conjunction with figure 4.2 can be used to view the locations of the highlighted generators in the network.

6.4.7 Reactive Power

Chapter 5 contained a study of generalized and general linear models with active and reactive power used as predictor variables. It was shown that the inclusion of reactive power variables in the model increased the fit of the GLM i.e. decreased the model deviance. In terms of logic regression, reactive power variables were added to the model to ascertain whether or not a lower model score could be achieved. Although the reactive power variables are not as strongly correlated to the damping, it is hoped that they may improve the model fit as the reactive power may further help account for some of the system dynamics.

In general, the reactive power outputs are not as well correlated with decay time constants as active power variables. However, reactive power outputs of machines give an indication of the power factors which can determine the levels of active power available to feed oscillatory instability. By including reactive power, interactions including active and reactive generation can be used to predict damping. As a result, the operating points of the generators can become useful tools in controlling the mode damping via power factor correction. In addition, it may be useful for operators to obtain information on reactive power support in order to formulate the best plan of action.

Figure 6.19 shows one of the most common interactions derived from the logic regression model with the reactive power variables included. The interactions as well as the variables have been tabulated as described earlier in section 6.3.2. The most common variables along with the most common interaction patterns are aggregated until the

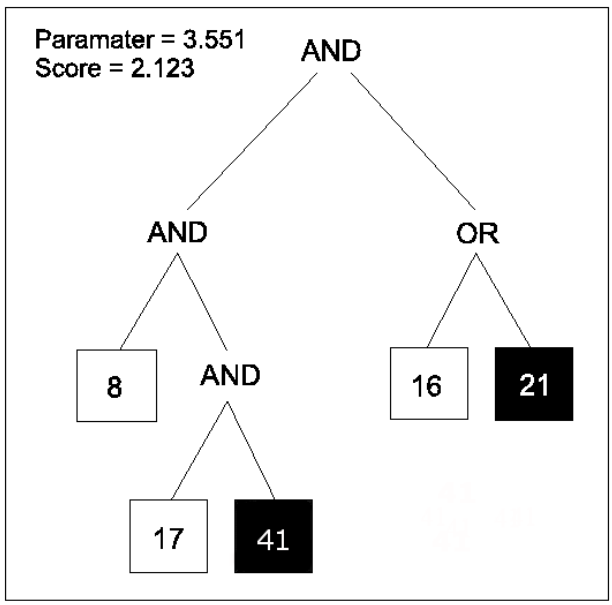


Figure 6.19: Optimal Logic Tree with reactive power variables included in the model

most common configurations are produced. These commonly occurring interactions can then be used to study the effect on damping even further.

Due to the increase in the number of predictor variables, the model size was increased from four to five leaves as the cross-validation step indicated that a five-leaf model was optimal. Figure 6.19 shows the new improved logic tree with a reduced model score of 2.123 and an increased logic regression parameter of 3.551. The model has retained variables 8 and 16 (GB27 and GB99) as expected but has a new AND pair of variables 17 and 41 which correspond to G17(GB102) MW and G18(GB104) MV variables. This is an interesting result as this model suggests that damping depends on a combination of mostly active power variables and a reactive power output.

In comparison to the average participation factors of figure 6.14 it can be seen that generator variables 8 and 16 are validated as these correspond to variables 11 and 24 in the participation plot (pre variable selection). Variable G17 (GB102) has also been validated using the participation factors as participation variable 25 (GB102 - Pre Variable Selection) was ranked fourth in the participation plot. This is in contrast to G14(GB52) which turned out to have a significant effect on mode damping but had a

relatively small participation factor.

Variable G17 has also been "anded" with variable 41 or G18 MVAr output. The predictor column variables were concatenated, with variables 1-34 related to MW variables and variables 35-68 being related to MVAr variables. Since both a MW and MVAr variables have been combined within an interaction, it will be interesting to see the effect this combination has on damping.
The G18(GB104) reactive power variable is interesting as it is negatively correlated with the damping i.e. decay time constant is high when G18(GB104) reactive power is low. This is interesting because G17(GB102) active power was reported in section 6.4.6 which is physically close to G18(GB104) and it highlighted MVAr output. Therefore, this combination may suggest that by changing the G17/G18(GB102/GB104) generation to increase reactive power for voltage support, a subsequent decrease in active power may help to reduce excessive decay time constants.

Table 6.6 contains the information from the MCMC step when it was applied with the reactive power included in the model. Variables G8 and G16 have more or less the same affiliations as before from the active power model. The two reactive power variables 40 and 41 (G17 and G18 or GB102,GB104) are also closely affiliated with the most common active power variables but also jointly appear in an appreciable number of models. These two variables are negatively correlated and hence have a (*) in table 6.6.

The interaction may not be a direct physical interaction as both active and reactive power are physically different quantities but the mode may be sensitive to both the active and reactive power generation.
In general, damping is improved when reactive power is increased which can be attributed a reduction in active power. Therefore, an intuitive grasp of the power system is required if a reactive power variable is reported as it will likely have a negative correlation with damping or may just be uncorrelated. As a result, reported reactive power variables may indicate regions sensitive to damping.

By including reactive power in the model, more information is made available to explain damping. This usually leads to a better model fit but can cause the significant β coefficients to become smudged over a larger number of coefficients. Figure 6.20 shows

MCMC Run		
Single	Double	Triple
8	21,16,17,23,14,41*	16,17,14,22,5
16	8,14,21,17,23,21	14,21,17,19,41*
41*	8,17,16,40*,19	17,21,23,16,11
40*	16,8,14,41*,15	14,8,41,17,11

Table 6.6: MCMC Comparisons from the most common active and reactive power variables

plots of the normalized G17 MW and G18 MVAr outputs superimposed on the 0.21Hz decay time constant record. These two variables formed an AND pair in figure 6.19. From the plot, it can be seen that the reactive power is at a relatively low value for the majority of the time when decay time constant is high. This translates as a high proportion of binary 0's for the G18 reactive power column in the logic regression model so there is a good chance it will be reported in the logic output as it will be negatively correlated with decay time constant. However, this doesn't necessarily mean it has an effect on damping.

The G17(GB102) MW output itself is not that well correlated with the decay time constant window and is included due to the fact that it is high during the poorly damped events. This is a common occurrence in the logic regression as high values that are assigned binary 1's tend to be significant in the model. The G18(GB104) reactive power output is sporadically correlated to the GB102 MW output but most of the time they are not coherent.

High levels of reactive power output from G18(GB104) may be due to loading conditions in that part of the network. High loading at the buses cause the voltages to drop and as a result reactive power is increased to support the voltage. These large loads may be responsible for dissipating some of the oscillation energy and reducing rotor swings in the generator.

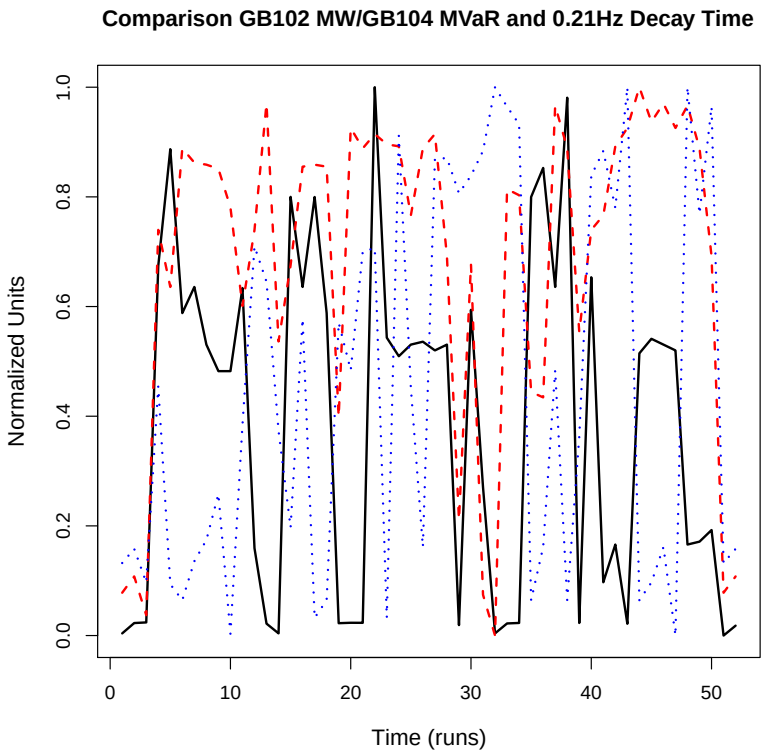


Figure 6.20: Comparison of 0.21Hz decay time constant versus G17(GB102) MW (red) and G18(GB104) MVAr (blue) outputs

The damping contribution of a generator is generally decreased as loading increases i.e. reduced rotor swings. This means that the increased reactive power output at GB104 should indicate adequate damping in that part of the network as oscillation energy will be absorbed by the high loading at the bus. However, the reactive power is flagged by the logic regression due to its negative correlation with decay time constant. G17(GB102) MW may have an effect on the mode damping but the reactive power predictor only indicates where there may be a voltage problem or a large load that may affect the damping. However, it may be such loads that are responsible for poor damping as they may draw power across a transmission corridor that sustain oscillations via inter-area rotor swings.

6.4.8 Development of Raw Logic Model

The development of the raw logic model takes place in a similar fashion to the wavelet transformed model, however there are some slight differences between them. First of all, the raw model presents less of a challenge in terms thresholding as the variance of the data is less than that of the wavelet transformed data.
The variable selection methods outlined in chapter 5 will be employed throughout this section so as to improve the logic model prediction. Once the raw logic model has been developed, sensitivity analysis will be used to determine the magnitude of the effect the reported variables have on damping.

6.4.9 Development of Logic Expression for Raw Icelandic Model

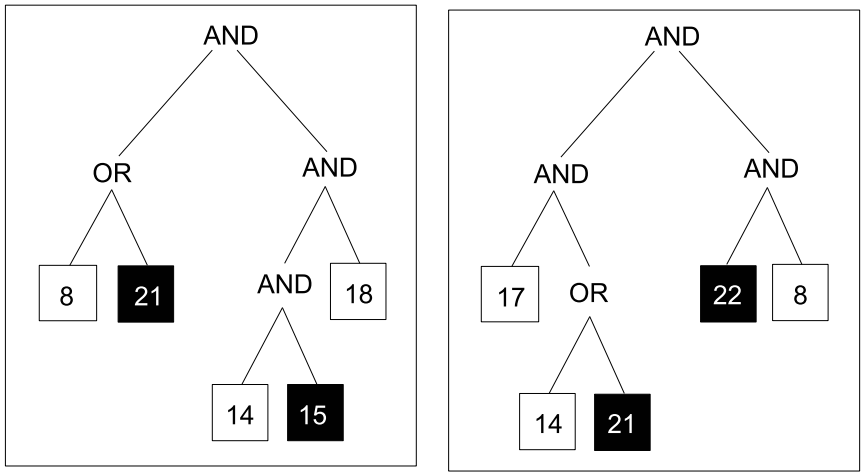
In this subsection, the 0.21Hz Icelandic mode was investigated using logic regression to determine the interactions leading to poorly damped events. The 256 length dataset was used in this study as this setup produced the best results in the wavelet transformed GLM. In addition, the longer dataset allowed a larger number of events to be used in the logic regression which may serve to improve the models. The raw data is more suited to a logistic logic model as the response/predictor relationship is non-linear. As a result, the entire raw dataset (response and predictors) was transformed into binary values and used in the logic regression.

Initially, active power variables were used to develop the model before reactive power variables were incorporated after. Variable selection has already been applied to the variables as explained in section 5.3.2, with the remaining variables used in this study. The key point in this section is the issue of highlighting the difference between the raw and wavelet transformed data outputs from the respective logic regressions. To follow on from this, both sets of results need to be reconciled and aggregated due to the fact that both sets contain valid observations that may be used to control mode damping.

Table 6.1 serves to highlight the slight differences between the raw and wavelet trans-

formed models which have also been compared to the participation factors to determine if the statistical models can be reconciled with the analytical model.

Figure 6.21 shows a comparison of logic trees derived from the initial models developed



(a) Five leaf logic tree output (b) Five leaf logic tree output

Figure 6.21: Comparison of outputs from raw logic regression step

with the 0.21Hz raw data. Subfigures 6.21(a) and 6.21(b) show two of the logic trees produced from the same thresholded data. Since logic regression relies on a number of algorithms to search through the state space there is always some variability in the logic functions that are produced. This is due to small differences in the selection and transition probabilities in the models that leads to a slight variation in the output.

However, the raw models are in good agreement with the wavelet transformed models covered earlier in this chapter in section 6.4.5. Familiar generator variables that were reported in subfigures 6.21(a) and 6.21(b) include variables 8,14,17,21 and 22 which correspond to generators connected to buses 27(8), 52(14), 102(17), 138(21)* and 139(22)* (GB138/139 are *not* generators*). See figures 4.1 and 4.2 for locational information regarding these generators.

From these logic trees, it can again be seen that there is partial agreement between the wavelet transformed logic models, the raw logic models and the wavelet transformed GLM from section 5.9.

However, the raw logic models presents different interactions in comparison to the wavelet transformed models. As well as this, some of the significant variables from the wavelet models are less common in the raw models and vice-versa. This subtle difference may have interesting consequences for the source location methodology as more variables are made available for testing and validation. It is this testing and validation that is investigated in the next subsection. By aggregating and combining both the raw and wavelet transformed results it is hoped that a more robust description of the system dynamics can be presented.

6.4.10 Reduction and Validation Methods

Reduction and validation methods were discussed in the previous section in an attempt to determine optimal model size as well as ranking the variables that make up the reported interactions.

Mutual information coefficients are especially suited to selecting significant raw variables as the mutual information was modelled on raw data. WGLM is not well suited to ranking the raw logic regression variables as the WGLM is based on wavelet transformed data as opposed to raw data. Again, tree reduction can be used to remove one leaf each time the model is run. In this way, the variables can be ranked in terms of importance in their specific logic tree.

Again, the cross-validation method (see section 6.4.4) was used on the raw data to determine the optimal model size. It was shown that the raw optimal model size was in fact a five leaf model as displayed in figure 6.22. This was further validated by the conditional permutation and greedy tests which also determined that a five variable model was the optimum size.

6.4.11 Results

In this section, results are presented from the raw logic regression based on the five leaf models derived from the cross-validation steps etc. The results were then tested against an eigenvalue analysis to see if there is a match between the models. Comparisons were also made between the reported predictors and decay time constants by superimposing

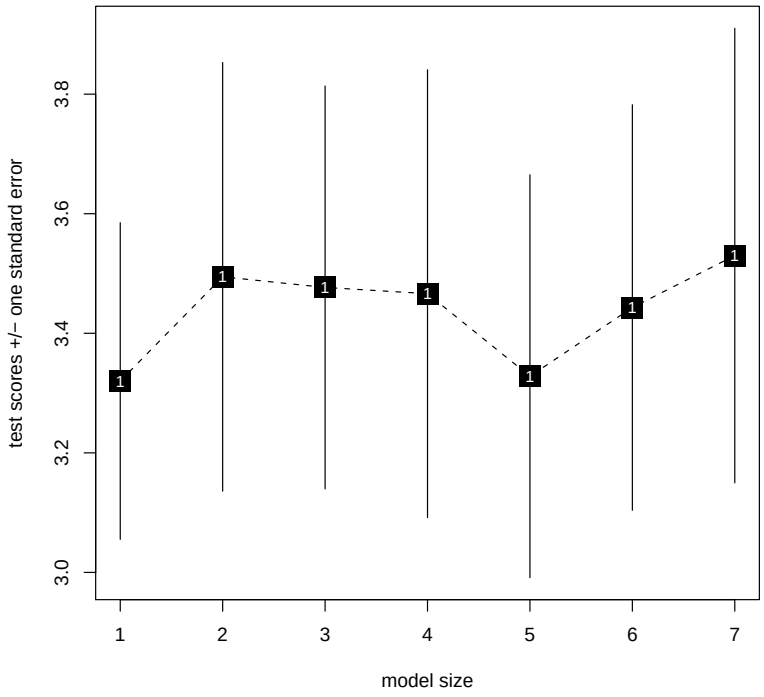


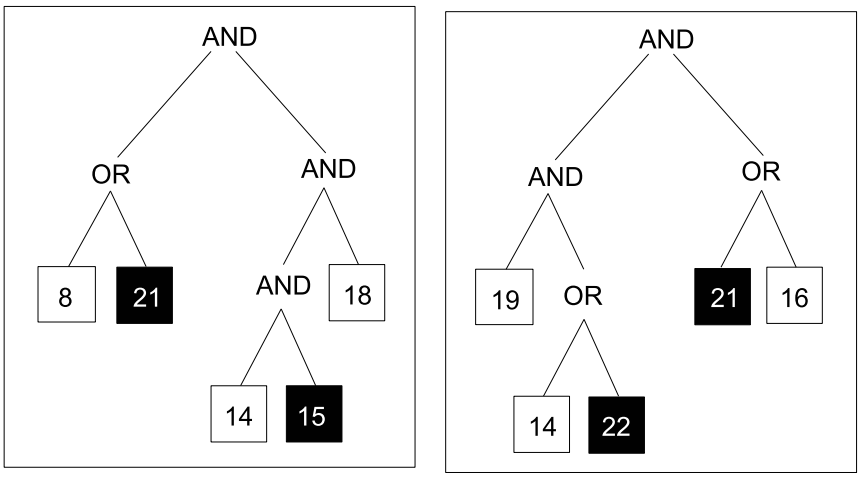
Figure 6.22: Cross-Validation step showing model scores for the number of leaves included in the model

their normalized graphs in a plot.

The results are derived from the Icelandic 0.21Hz mode which has already been subjected to variable selection and pre-processing which leaves a total of 23 generators used in this study, with the slack bus also removed.

Figure 6.23 shows a comparison of two five leaf models derived from the Icelandic 0.21Hz logic regression model. It can be seen that the outputs vary quite a bit with a total of six pro-active variables being represented between both models, which is similar to the wavelet transformed case. The MCMC step allows all the variables in each model to be tabulated and compared against each other to determine how many times they appear in the same model. This in turn allows us to establish the most common interactions produced within the model as well as removing some of the redundant variables and interactions.

Table 6.7 shows the comparison of the number of observations of joint variables in the MCMC raw models. The single column shows a selection of six common variables in the MCMC step with generator variable 21 (GB138) being the most common. The



(a) Five leaf logic tree output (b) Five leaf logic tree output

Figure 6.23: Comparison of outputs from raw logic regression step in terms of generator number

MCMC Run		
Single	Double	Triple
21	8,14,1,18,15	14,16,18,1,15
8	21,14,16,19,15	14,16,18,1,15
14	8,16,21,18,5,	21,16,15,18,19
16	14,19,3,21,10	8,19,3,17,15
18	14,21,8,5,19	21,16,8,10,21
15	18,8,14,19,21	8,21,19,14,17

Table 6.7: MCMC Comparisons from the most common generator variables derived from the raw models

double column displays the most common variables that appear in the models alongside the single variable in column 1. For example, the double column in row 2 contains the variables 21,14,16,19,15 which means that variable G21 (GB138) is jointly observed with variable G8 (GB27) across the state space more than any other variable. Variable G14 (GB52) is the next most common variable jointly observed with variable 8 an so on.

The triple column was explained earlier in this chapter but shall be repeated here for convenience. The triple column contains the most frequently observed variables in the MCMC models that also contain the single variable and the most common double vari-

able. For example, the row 3 triple column contains the variables 21,16,15,18,19. This means that these variables appear more commonly in models containing variables G14 and G8 than any other variable, with variable G21 being the most common.

The MCMC single raw variables seem to be in good agreement with both the wavelet transformed logic regression results shown in table 6.3 as well as the results from the wavelet transformed GLM in chapter 5 section 5.8.3 with G8(GB27), G14(GB52) and G16(GB99) being reported as significant. A large number of the common single variables also appear jointly in the model with other common single variables as evidenced by the double column. For example, variable 8 (GB27) also appears quite regularly with variables 21 and 14 (GB138 and GB52), which are also common single variables. In contrast, it can be seen that variable G8 is also jointly observed with variable G19 (GB117) even though variable G19 is not in itself all that common.

This is a valid point and should be highlighted as a key difference between the raw and wavelet transformed models. Variable G19, as well as variable G16, were common variables in the WT logic models that were validated via the MCMC step. Variable G16 (GB99) in particular was shown to be a critical variable in terms of mode decay time constants as by reducing it caused the decay time constants to be reduced across a large number of mini-events. In the raw case, variable G16 is ranked fourth most common as opposed to second for the WT model. This is why both models are used in the source location methodology as both models extract different sensitivities in the system. Variable G18 is ranked fifth in the raw model, so it should be interesting to see the effect this variable has on mode damping.

Variable G19 (GB117), which was ranked fifth in the WT model is only reported in the raw model due to its numerous appearances alongside more common single variables. Again, the importance of this technique can be seen as significant variables may not occur all that commonly on their own but may always occur in conjunction with pairs of significant variables. Therefore, if the significant variables are validated then other variables that occur in conjunction with them in the logic models must be considered

as well.

The point of the MCMC step is to highlight commonly occurring groups of variables across a number of models. This allows the less frequent variables to be included in the model by virtue of association with common variables. This can be seen with variable G19 (GB117) as it hasn't been reported as an important single variable i.e. its not in the top six, but is present in quite a few models containing significant variables. As we will see later, these less common values can have a significant influence over the damping in combination with certain other variables.

However, the emphasis is placed mainly on common single variables and their interactions with other common single variables. This was seen in section 6.4.6 where G8(GB27) and G14(GB52) produced an interaction that seemed to induce poorly damped periods.
By using these double and triple values, results can be obtained across all the possible models which means that no information is lost as may be the case with running just a single model.
The use of this aggregate method spanning a large number of models is ideal for the source location problem due to the large number of interactions that take place in the power system. Although, some variables are more common than others, it is their interactive effects that matter and as a result no significant variable produced from the MCMC step (single, double or triple) can be discounted as they may prove to be very significant in terms of mode damping.

6.4.12 Reliability of Raw Icelandic Logic Model

The reliability of the raw Icelandic model used in this work will be tested and validated to determine the accuracy of the model. As in chapter 5, participation factors derived form eigenvalue analysis will be used to compare the logic regression results with the dynamic performance of the system to see if there is any comparisons to be drawn.
In addition, plots of the significant variables from the logic model will be compared with the mode decay time constant to determine the levels of correlation present. Ad-

ditionally, since the logic model focuses on the interaction effects, the dynamic model will be modulated to determine the extent to which these interactions affect the 0.21Hz mode.

Figure 6.24 shows a typical logic regression tree from the raw Icelandic model. This

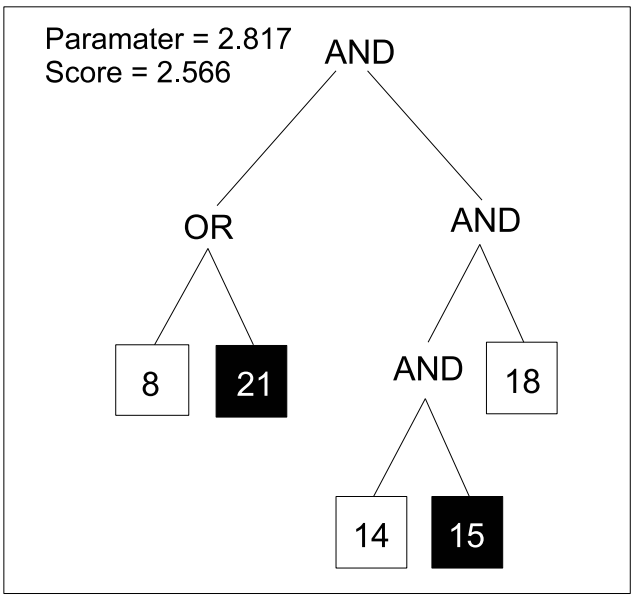


Figure 6.24: Optimal Logic Tree from Icelandic Model with Score and Parameter Values

model was presented earlier in figures 6.21 and 6.23 as it is one of the most common interactions produced by the raw logic model.

As previously mentioned, the logic models are subject to some variation and as a result, the reported variables may differ from run to run. In the previous sections, methods were discussed to overcome this difficulty with the use of Monte Carlo simulation to determine common groups of variables. It must be noted that the MCMC step is used to supplement the logic models.

The raw Icelandic data lends itself to quite a well fitted logic model. However, the model is not as well fitted as the wavelet transformed case which is to be expected due to the optimal scale components of the wavelet transform. Although the fit isn't as good, there is still the possibility that the wavelet transformed model may omit some

information that may be retained in the raw model. As a result, the raw model is used in conjunction with the WT model.

However, the raw model is still quite well fitted and the variation in the outputs from run to run is minimal. Some variation does exist and as a result, aggregate methods are still required. Once again, variables G8, G14 and G21 (GB27, GB52 and GB138) appear in the logic tree in figure 6.24. These variables were shown to be significant in the MCMC step as they appeared in many of the Monte Carlo models.
Variables G15 and G18 (GB53 and GB104) also appear in the tree. Both these variables are shown to be common in the models via the MCMC step, although they are not present in the wavelet transformed models. As a result, these variables are a good starting point in determining the predictors and interactions that are strongly linked to mode damping.

In addition, by investigating these different variables it is possible to determine the differences in the raw and wavelet transformed models which allows us to reconcile and aggregate the models appropriately. It is hoped that the raw model may contain some sensitivities not highlighted in the WT model and vice versa as both models overlap quite significantly.

Table 6.8 exhibits the MCMC results for the variables reported in the initial logic

MCMC Run		
Single	Double	Triple
21	8,14,1,18,15	14,16,18,1,15
8	21,14,16,19,15	14,16,18,1,15
14	8,16,21,18,5,	21,16,15,18,19
18	14,21,8,5,19	21,16,8,10,21
15	18,8,14,19,21	8,21,19,14,17

Table 6.8: MCMC Comparisons from the variables derived from a common raw logic model

regression results shown in figure 6.24. The point of this exercise is to compare the logic trees and their affiliated variables with the dynamic participation factors to see if there are any matches. Additionally, the interactions will be tested in the dynamic model

by examining their effect on mode damping when some of the interaction variables are modified.

Figure 6.25 displays the average participation factors for the Icelandic 0.21Hz mode

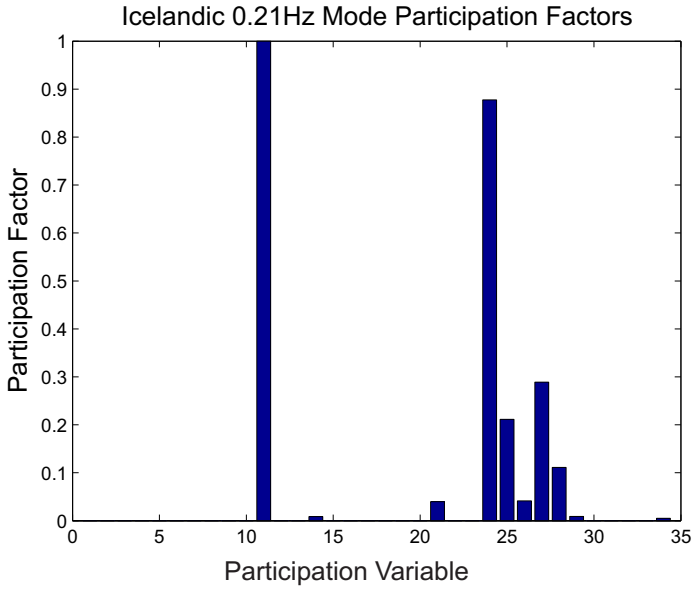


Figure 6.25: Icelandic 0.21Hz Mode Participation Factors derived from rotor angle states

under consideration. The results have been normalized with the largest value attributed to variable 11 which corresponds to G8 (GB27). These have been shown again here for convenience and are exactly the same as the participation factors in figure 6.14 as both the raw and wavelet transformed results are compared to the same analytical model. It must be noted that the 34 variables in the plot represent the 34 generators in the model with the slack generator removed i.e. the original model had 35 generators. In addition, variable selection was applied to these original 34 generators resulting in the new index used in the WGLM and logic regression models.

As expected, there is some agreement between the participation factors and the logic regression step just like the wavelet transformed case. Again, variables 11 and 24 (G8(GB27) and G19(GB99)) are clearly the highest participating machines in the system. Other machines that have a higher than average participation include G17(GB102) and G19(GB117) as well as G14(GB52). In addition, variable 26 or G18(GB104) has

a slightly elevated participation factor as well but like GB52, it is much smaller in comparison to the largest participation factors.

However, some of the commonly reported generators from the raw value MCMC step have negligible participation factors relative to the highest participating machines. The focus of this section is to determine the effect these machines have on the damping in conjunction with the more obvious generators. This must be studied as there is some disrepancy between the logic regression step and the participation factors, the source of which may lie in the interaction effects of the generators.

In the previous logic regression analysis we presented a number of interactions that looked like they had an effect on the mode damping. Together with the cross-validation and MCMC steps it is possible to produce a list of variables in order of their significance. This rank is based on the number of times they appear in the model as well as the number of appearances in conjunction with significant (single column) variables.

Table 6.9 shows the comparison between the aggregated logic regression variables and

MCMC Run		
ModeFreq	LogicReg	PF's
0.21Hz Gens.	G21,G8,G14,G16,G18,B53,	G8,G16,G19,G17,G20,G18
0.21Hz Buses.	138,27,52,99,104,B53,	27,99,117,102,126,104
0.21Hz Gens.	G8,G14,G16,G18,B32,G19,	G8,G16,G19,G17,G20,G18
0.21Hz Buses.	27,52,99,104,B32,117,	27,99,117,102,126,104

Table 6.9: Significant variable comparison between logic regression and participation factor results in terms of generator number and generator buses

the participation factors for the 0.21Hz mode. The first row of results contains negative Boolean variables i.e. G21(GB138) is a NOT value, whereas rows 3 and 4 contains the ranked logic variables with the NOT values removed. It can be seen that both sets are generally in agreement except for the presence of G14(GB52) in the logic analysis which isn't present in the participation factors.

It is interesting to note that a negative or NOT Boolean variable is actually the most common variable in the model. This suggests that the G21(GB138) output is not a

critical factor in the 0.21Hz mode as its high output periods are generally not correlated with high decay time constants. In general, generators that are actively high during poorly damped periods are better suited for controlling modes.

Referring to figure 6.24 which shows one of the most frequent interactions produced by the logic regression step, it can be seen that variable G14 (GB52) is part of a nested AND statement and should have a direct effect on damping. The same can be said for variable G18 (GB104) which wasn't reported in the wavelet transformed model. This variable is part of a nested AND statement also and will need to be investigated.
In fact, from the sample runs it was shown that whenever variable G18 appears in a model with variables G8 or G14 they are always connected via the AND function which means either G8 AND G18 OR G14 AND G18 OR a combination of both combine to create the conditions for poor damping.
In section 6.4.6 it was shown how generators at generator buses 27(G8) and 99(G16) contributed most to 0.21Hz events as they were prioritized in the logic models as well as the MCMC regression. The ranking process also identified variable G14 (GB52) as an important factor even though it had a relatively low participation factor. This variable was subsequently validated by modulating generator G14 at generator bus GB52 to produce decay time constant profiles which were partially degraded in relation to the original plots (see section 6.4.6).

In the raw model, variable G18 (GB104) serves this purpose. Again, variable G8 is the most significant pro-active variable and variable G14 has also been reported as significant. Generator variable G16 (GB99) is reported as well although it is more significant in the WT logic model. Since three out of four of the top pro-active variables in table 6.9 have been validated in the WT model, it would be useful for the remaining variable (variable G18) to be investigated to determine if it has an effect on damping. If this proves to be the case, then both the wavelet transformed and raw models can be used together to give a more complete description of the causes of poor damping in the system.

Figure 6.26 shows the selected decay time constant profiles when both G8(GB27) and G18(GB104) are replaced simultaneously with a low constant power source. Both were reduced due to the aggregate AND operator which was shown to connect both variables in the sample models.
Again, this plot is quite revealing as the initial "mini-event" from 4-10 seconds has been diminished relative to figure 6.17. In addition to this, the mini-events at 34 and 36 seconds are no longer present in the subsequent loadflow runs even though they can still be partially observed when only G8(GB27) is modulated (figure 6.17).

This suggests that the interaction of G8(GB27) and G18(GB104) has an effect on mode damping as suggested by the lowest scoring model tree in figure 6.24. However, G18(GB104) is not as critical as G14(GB52) from the WT model as the decay time constants, when G18(GB104) is reduced, are only partially diminished in comparison to the G14(GB52) case in figure 6.18.
Since these mini-events were partially diminished when both machines (G8(GB27) and G18(GB104)) are reduced, this means that both these machines can be used to provide a more robust control of damping.

In section 6.4.7, results were presented that indicated that G18(GB104) reactive power was in some way correlated or significant in terms of mode damping. In that example, the correlation was negative in that a high reactive power output at G18(GB104) meant that the mode was well damped. It was proposed that high loading in that area caused an increase in reactive power and as such any oscillation would not be related to G17/G18 (GB102/GB104) active power as the load would dissipate the oscillation energy.

However, since G18(GB104) active power generation has been reported in the raw logic regression it must be the case that active generation at G18 can affect mode damping. Therefore, any reactive power support from G18(GB104) must be sent across a transmission line to a high loading area elsewhere in the network and is not related to high

loading at bus G18(GB104), otherwise any load here would absorb energy from GB104 and thus not produce oscillatory behaviour.

It has been shown in this case that generator G18 at GB104 has some effect on

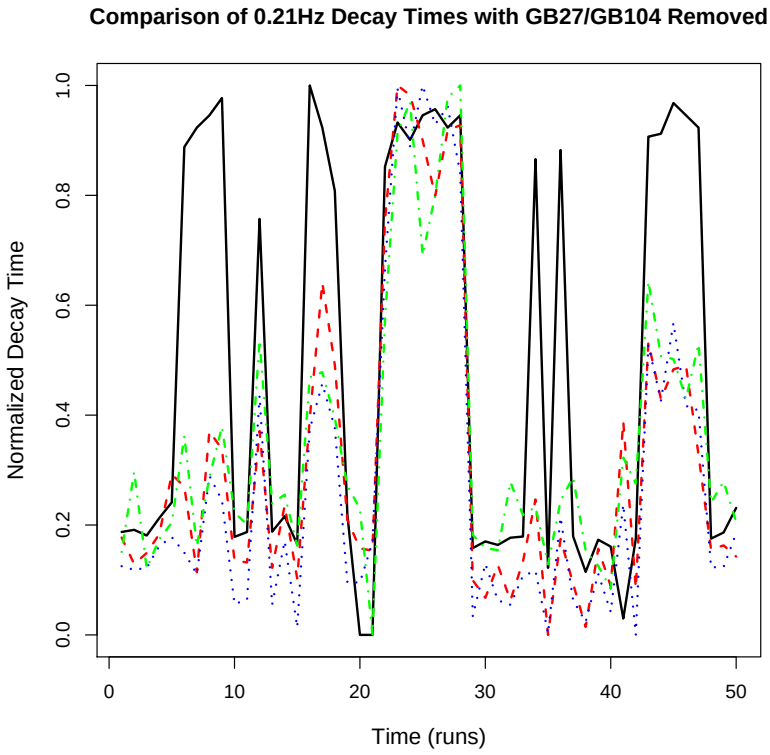


Figure 6.26: Comparison of selected decay time constant plots when G8(GB27) and G18(GB104) are modulated

the 0.21Hz mode damping. It was shown that G18(GB104) has a similar effect on the mode damping to that of G14(GB52) which was highlighted in the wavelet transformed model. This is an interesting result as both models have an exclusive variable that interact with more significant variables to diminish damping. This exclusivity principle warrants the use of both the wavelet transformed logic model and the raw model as both models produce pieces of exclusive information that can be used to control damping. This in turn allows the mode damping to be more robustly controlled as a greater number of significant variables are made available.

This example has displayed the power of logic regression. It has been shown that

G8(GB27) is a significant variable in terms of mode damping. In addition, it was shown how G14(GB52) or G18(GB104) also contribute to damping in conjunction with G8(GB27) as if either of them are low while G8(GB27) is low then a number of mini-events become diminished. This means that as many as three generators can be used to control of mode damping.

There is a considerable amount of overlap between the raw and wavelet transformed logic regression models. Seeing as the interaction effects evolve over time, it can be difficult for a given logic model to encompass all sensitivities taking place in the system. In the 0.21Hz mode case, the wavelet transformed model performed better as it produced more pertinent interactions and variables. However, the raw model reported some variables and interactions not evident in the WT model. The model score depends on the how well the data for a particular mode is fitted to a given regression. In this case, the WT model performed better but this may not be the case with different sets of data.

The numerous interaction effects derived from the raw logic model will be explored in more detail in chapter 7. In this chapter the initial interactions and those variables derived from the MCMC steps were shown to be quite significant. These examples have shown the source location problem to be more complex than originally thought. This is due to the large number of interactions responsible for poor damping as opposed to the effects of one or two system variables.

6.4.13 Reactive Power

In section 6.4.7, the performance of the logic regression model was investigated when reactive power was used in the model. It was shown how the model score was improved with the reactive power addition as well as the inclusion of a number of reactive power variables in the subsequent logic tress and Monte Carlo Markov Chain regressions.

In general, reactive power outputs are not as well correlated with decay time constants as the active power outputs. However, reactive power generation gives an indication

of the operating point of those machines which can be used survey the importance of a generator in terms of mode damping. By including reactive power, interactions including active and reactive flows can be used to predict damping. As a result, the operating points of the generators can become useful tools in correctly controlling as well as adjusting MW and MVAr outputs and thus mode damping.
The addition of reactive power variables to the raw model has again caused the optimum tree size to increase by one, from five to six leaves. This is also true of the WT model which increased from four to five leaves. Once again, the addition of the reactive power variables served to improve the model score as it was reduced from 2.566 to 2.421 while the logic regression parameter increased from 2.817 to 3.053 on average.

Table 6.10 contains the information from the MCMC step when it was applied with the reactive power included in the model. Just like the WT models, all the significant active power variables have been retained. These include variables G21 and G8 which have more or less the same affiliations as before in the active power model.
The notable feature in table 6.10 is the fact that the reactive power variables 40 and 41 which refer to G17 and G18 (GB102 MV,GB104 MV) have been reported again via the raw MCMC. This is a nice result as variable G18 (GB104 MW) was reported previously in the raw logic model. The fact that the reactive power variables have some correlation with the decay time constant records shows that they contain some information on damping.

The two reactive power variables 40 and 41 (G17,G18(GB102,GB104)) are also closely affiliated with the most common active power variables as they jointly appear with them in an appreciable number of models. This means that the reactive power output in this section of the system may be linked to the damping even though the reactive power variables are NOT variables which suggests a negative correlation. However, it was discussed in the previous section how the operating point of G17,G18(GB102/GB104) generators may affect electromechanical mode and as such, reactive power variables can be used to determine a suitable course of action to control damping.

MCMC Run		
Single	Double	Triple
21	8,18,1,14,41*	18,5,14,41*,19
8	21,14,19,40*,16	18,5,14,41*,19
41*	21,40*,8,18,19	8,40*,5,14,11
40*	14,8,41*,18,16	41*,8,19,21,22

Table 6.10: MCMC Comparisons from the most common active and reactive power variables

6.5 Summary

This chapter introduced the concept of logic regression as a method of establishing important interactions in the power system that make poor damping more likely. The focus of the chapter was on the 0.21Hz Icelandic mode that was investigated in chapter 5 using the general linear model. Both wavelet transformed and raw data were used to determine which produced the best predictions with minimum error.

Logic regression models were initially developed from wavelet transformed data with variable selection and optimal wavelet scales applied. The resulting logic models derived from the Monte Carlo Markov Chain regression were used to determine the most common interactions that occurred in the system and how they led to poor damping. G8(GB27), G16(GB99) and G14(GB52) were shown to produce a significant interaction that was linked to the poorly damped events.

The wavelet transformed logic models agreed quite well with the WGLM and participation factors from chapter 5, though some of the significant variables from the logic regression had negligible participation factors. This was the case with G14(GB52). This indicated that there were some subtle differences between the results of the statistical and analytical models. Reactive power was then added to the logic models and it was noted that the model score decreased from 2.301 to 2.123, which indicated a better fitted model.

The raw data was then used to develop interactive logic models. The result was very

similar to the wavelet transformed case in that G8(GB27) and G16(GB99) were reported as the most significant. Interestingly, G18(GB104) was highlighted as an important variable in the logic regression study, but again, it had a negligible participation factor. By reducing the magnitude of combinations of generator outputs including G18(GB104), it was shown that G18(GB104) did have an effect on mode damping and that the differences between the statistical and analytical models were justified. The validation of these reported interactions showed that certain generation patterns in the system were conducive to poor damping.

The difference between the raw and wavelet models was highlighted by exclusive variables that were only reported in either of the models but not both. In the raw case, a five leaf model was shown to be optimal and a best scoring model of 2.566 was presented with a logic regression coefficient of 2.819. G18(GB104) was highlighted as an important variable with a small participation factor, whereas it was not reported in the WT model. In essence, both models produced significant variables that weren't reported in the other model. In the case of G14(GB52), it was reported by both models even though it has a negligible participation factor.

Again, reactive power was added to the raw logic models which had the effect of decreasing the model scores. In addition, G17,G18 (GB102/GB104) MVAr variables were reported in the subsequent logic regression analysis which was an interesting result as GB104 MW was also highlighted in the active and reactive power models. The interaction of both active and reactive power in certain sections of the grid allowed a more detailed explanation of the mode dynamics to be developed.

Chapter 7

Case Studies using Source Location

7.1 Introduction

In chapters 5 and 6 it was shown how general linear models and logic regression could be used to determine significant variables and interactions in the power system that can make poor damping more likely. Various methods were introduced to further improve the model fits in an attempt to make the model predictions more accurate and thus improve the accuracy of the significant predictors. A number of examples were used to help explain the various procedures used in the methodology. The examples consisted of the 16 machine model and the Icelandic power system model as these were large enough to allow the WGLM and logic regression models to be developed and compared to the dynamic performances of the modelled systems.

In this chapter, all the components presented in chapters 5 and 6 are brought together with the aid of two case studies. The first case study is the 0.62Hz interarea mode from the Icelandic model. The 0.62Hz mode allows the logic regression to be taken a bit further as it is more complex than the 0.21Hz mode introduced in chapters 5 and 6. By validating the methodology again in terms of this more complex mode it further validates the technique, which is essential if the methodology is to be implemented into existing software for use in a real power system.

The second case study used in this chapter is an analysis of the Australian power system with real measurements derived from various system buses and inter-ties. The Australian system is a well known network with a history of some poorly damped events. The mode of interest here is the 0.48Hz interarea mode that exhibits maximum decay time constants of 60-70 seconds which gives an indication of some of the stability problems experienced in the system. By validating the algorithm using model-based studies, a system operator can have confidence in the output from the source location methodology as opposed to a methodology that has not been robustly tested.

In the case of the real system, the results can only be validated through surveying model fits and by superimposing the results against the decay time constant records. If PMU penetration was more comprehensive, participation factors could be deduced from the real system and compared to the statistical results but this cannot be implemented due to low PMU penetration.

The validation process for the modelled data would consist of surveying the statistical model fit to determine if it was acceptable. If so, the significant loadflow variables highlighted by their regression coefficients could then be adjusted to see if they affect the corresponding mode decay time constants. If so, then the statistical model can be considered to be well validated. This does not automatically mean that the source location technique is fully validated for real power data, as this can only be achieved through extensive testing in real time with data from a real system.

Again, the Icelandic dynamic model will be modulated to determine the effect certain interactions have on the 0.62Hz mode damping and to show the effectiveness of the methodology. Again, this cannot be done in the real system but it is hoped by further validating a new system mode, that the implementation of a real source location system could be made possible in light of the model validation process. If this is the case, then the reported interactions and predictors could be used as a means to control mode damping [2].

7.2 Case Study: Icelandic 0.62Hz Interarea Mode

7.2.1 Icelandic 0.62Hz Interarea Mode

The 0.62Hz Icelandic mode has been used in this case study due to the excessive decay time constants observed at this mode frequency. Figure 7.1 shows the spread of eigenvalues derived from the model setup where random bus injections and intentional generator ramps were used to simulate the system behaviour. The observability and participation factors, which are detailed later, have been used again to validate the mode as there is quite a lot of movement in the frequency as the system conditions change. This interarea mode is one of the most interesting modes in the system as it

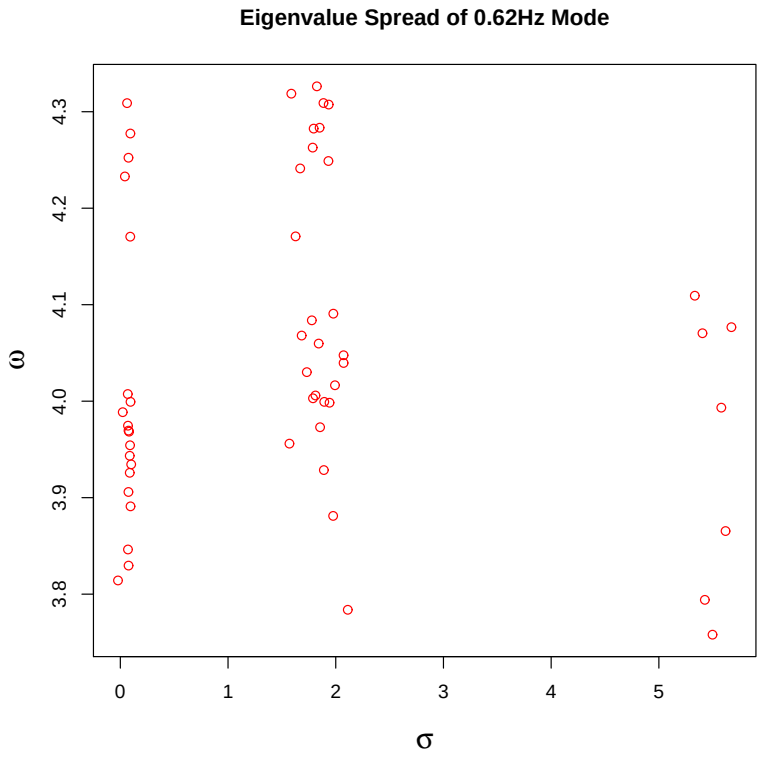


Figure 7.1: EigenValue plot of 0.62Hz interarea mode

involves a combination of generators located throughout the network and has exhibited decay time constants in excess of 40 seconds. The 0.62Hz is suited to the methodology as it has all the qualities required for a good testing and validation procedure.

7.2.2 Development of 0.62Hz Wavelet Transformed Generalized Linear Model

Initially, a wavelet transformed general linear model (WGLM) was developed for the 0.62Hz mode before logic regression was applied to the data. The model was setup with the random permutations and intentional adjustments so as to induce some poorly damped events. In this case, window sizes of 128 and 256 were used as these were of sufficient length to allow enough events to occur so that a reliable statistical model could be developed.

A maximum data length of 256 has been imposed in this study as using too long a window may serve to clutter the model and reduce the prediction accuracy. This is especially true with power systems because they tend to evolve with time, as different combinations of variables begin to effect damping. By spreading a window over a long time, the regression coefficients can become smudged and it can be difficult to interpret the results. This serves to inhibit the prediction capability of the model and its ability to identify strongly correlated predictors.
This is normally the case with the dynamic models as simulations tend to be set up for relatively short durations to reduce computational times. However, real systems may be stable for much longer and as a result, longer data windows can be used in such analyzes.

The aim of this section is to develop a GLM from wavelet transformed variables to yield a model that explains the dynamic behaviour of the system. It has been observed that the signals contain redundancy i.e. some frequency bands contain little or no interesting dynamics. By using optimal scales of wavelet frequency bands it is hoped that the model can be further improved to enhance predictability as well as preserving and highlighting significant predictors that can be used to influence damping.

7.2.3 Variable Selection

In section 5.6.3 we discussed how some preliminary thresholding can be done on the system state variables to simplify the resulting GLM. Table 7.1 below contains information on four of the variables removed in the preliminary stage. Both the mean and the variance have been shown for the 256 dataset model. It can be seen that both the means and the variance are relatively low with the variances corresponding to maximum swings of 1 to 1.5MW from the mean (see figure 5.8).

Since these generators are supplying negligible oscillation energy to the mode in question and because they do not swing enough to seriously perturb the system they have been omitted from the model. In this case, the mean threshold has been increased to 7MW in relation to the longer decay time constants seen for the 0.62Hz mode. This variable selection allows the larger generators or groups of generators to be used exclusively in the statistical model. This will serve to highlight the truly significant loadflow variables further as the smaller less significant variables will not be present to smudge the results. In section 5.3.2 we discussed two methods of variable selection

Variable	Definition	Mean	Variance
26	Gen Bus 104	5.74	0.33
27	Gen Bus 117	5.11	0.39
32	Gen Bus 142	3.12	0.31
34	Gen Bus 186	6.04	0.42

Table 7.1: Mean and Variance values for predictors removed in preliminary thresholding

that allowed the removal insignificant and multicollinear variables from the statistical models. In this section both mutual information and PLSR has been applied to the Icelandic 0.62Hz data to determine if any redundant variables can be excluded. Table

Mode	PLSR	M.I
0.62Hz	4,14	13,28,51

Table 7.2: Generator Buses removed from model via PLSR and mutual information steps

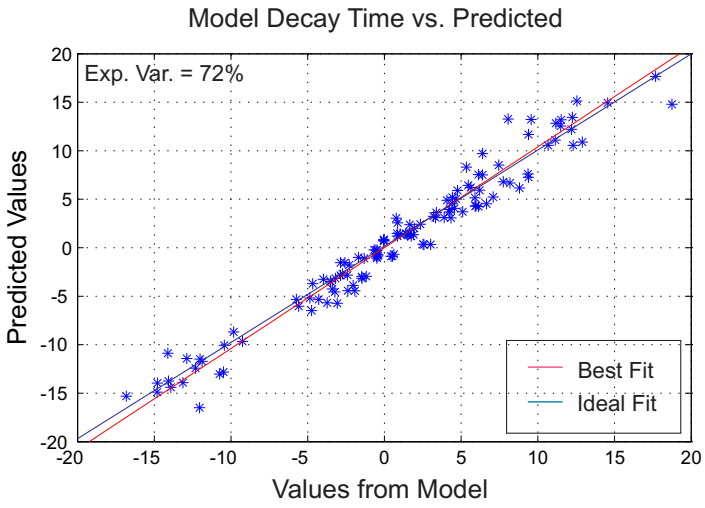
7.2 shows the generator bus variables removed, via the mutual information and PLSR models, for the Icelandic model data with 256 data points. Generator buses 13,28

and 51 have been excluded as they are deemed to have negligible mutual information with the response signal (decay time constant). Generator buses 4 and 14 have been excluded as the PLSR model has deemed them to have a small effect on damping in relation to the PLSR latent vectors. As a result, the variables in question cannot be the true underlying cause of response variability. In the case of the redundant PLSR variables, they may be correlated with a true predictor but are not the cause of poor damping as they do not account for much Y-space variance in relation to the highest scoring latent vectors. Colinear variables may be reintroduced into the analysis after the variable selection if their collinearity has physical meaning in relation to the significant predictors.

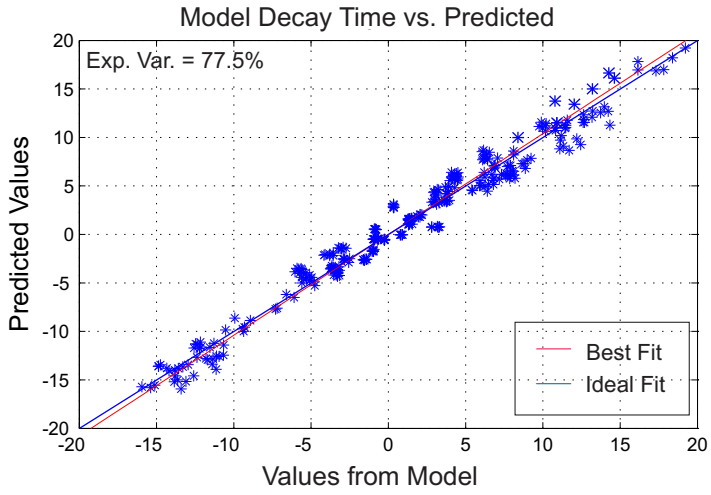
Although both the mutual information and PLSR don't always reconcile, it is useful to combine them to remove redundancy from the model. The MI step gives us an idea of information shared by two signals whereas the PLSR allows us to determine which variables have a negligible effect on the response variable in terms of their interaction with other predictors i.e. PLSR determines an optimum set of predictors to explain Y variance where collinear variables serve to reduce the latent vectors scores. However, care has to be taken to ensure that too many variables aren't removed as this could inhibit the model prediction capability by the removal of genuinely significant predictors.

7.2.4 WGLM Model Fit

WGLM's were constructed using both decay time constant and system state datasets as outlined in chapter 5 section 5.5.2. The subfigures in figure 7.2 show the plots of the actual wavelet coefficients derived from the decay time constant versus the predicted values derived from the GLM. These plots are shown for the same optimal scale combination with two data lengths, namely 128 and 256 points. A best fit line has been added to the plots to show any linearity that indicates a well fitted model. Referring to figure 7.3 it is clear that most of the dynamics of the signal occur at the third, sixth and seventh resolution levels for the 256 length data. Even though there are a few sporadic and relatively large coefficients at the redundant levels, these signal resolution



(a) GLM Prediction with 128 data length and optimal scales



(b) GLM Prediction with 256 data length and optimal scales

Figure 7.2: Modelled Wavelet Coefficients from power system model versus Predicted Wavelet Coefficients

levels encompass spectra that do not represent the bulk of the dynamic content of the signal.

In subfigure 7.2(a) the 128 data length scenario is presented. In this instance, only

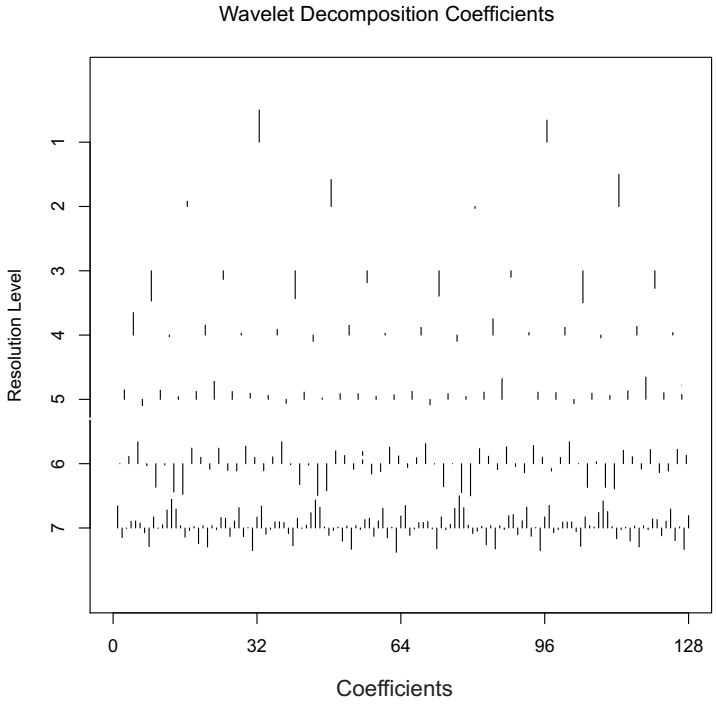


Figure 7.3: Wavelet Coefficients at various resolution levels derived from Icelandic 0.62Hz mode decay time constant

resolution levels 2,5 and 6 are used in the GLM as these resolution levels contain a large majority of the dynamics. Levels 2,5 and 6 correspond to levels 3,6 and 7 in the 256 length data as more resolution levels are added each time the data length is doubled.

For 128 length data ($f_b = 1/64$), resolution levels two, five and six correspond to discrete frequencies 0.0625Hz, 0.5Hz and 1Hz (highest resolution) respectively where 1Hz corresponds to a coefficient describing changes in the signal occurring between each sample i.e. 64 coefficients for 128 length data.
For 256 length ($f_b = 1/128$), resolution levels three, six and seven correspond to discrete frequencies $8 * f_b = 0.0625$Hz, $64 * f_b = 0.5$Hz and $128 * f_b = 1$Hz respectively where $128 * f_b = 1$Hz again corresponds to a wavelet coefficient describing changes that occur

in the signal between sample periods (256 in total) i.e. 128 coefficients. Therefore, for double the length of data set, each frequency band is just shifted up a level.

The wavelet coefficients determine the relative contribution to a signal of each frequency component at a given point in time i.e. levels 3 and 4 coefficients determine how much of the raw signal is composed of 0.125Hz and 0.25Hz dynamics, for 128 data length. It can be seen that when the redundant frequency bands of the signal are removed from the GLM the prediction is greatly improved. In this case the null deviance is 2501.5 with 104 degrees of freedom with the residual deviance at 700.4. As a result 72% of the variation is explained in the model.

Subfigure 7.2(b) again shows a plot of the wavelet transformed actual decay time constants versus the estimated decay time constant. In this case a dataset of size 256 was used to see the effect it would have on the wavelet transformed model in comparison to the 128 length case. A longer data set allows more of the high resolution detail to be used in the GLM. In this case resolution levels 3,6 and 7 are used where 6 and 7 correspond to the highest resolution levels. It must be noted that by using the optimal scales the number of data points is reduced. This is due to the optimal scale levels being retained whilst the others are removed. In this case, levels 3,6 and 7 correspond to 8, 64 and 128 data points which means 200 values are used in the WGLM.

In this case, the 256 length model explains 77.5% of the response variance which is a better result. The null deviance is 4322.3 with 208 degrees of freedom with a residual deviance of 972.2. Once again, an improvement in the model is seen when the redundant data is removed and more concentrated dynamic data included in the GLM.
The results show that significant improvements can be achieved in the GLM if the optimal wavelet scale combinations are used in the analysis. Depending on the characteristics of the signal, longer data sets can be used to extract more finer detail from the signal in terms of both time and frequency. By determining which combinations of wavelet resolution levels reduce the overall deviance of the model, it is possible to accurately determine the significant predictors that affect mode damping.

In summary, a larger dataset allows more higher resolution data points to be used in the analysis. Depending on the signal characteristics this may produce a more accurate wavelet transformed GLM as a greater number of data points can reduce the model error. However, due to the dynamic nature of the power system, using too long a dataset can result in some important information being lost.

In this instance, the true predictors can become less meaningful due to the number of interactions taking place over the longer window. In general, less complex systems need shorter data sets to produce a good fit as the relationships are usually more obvious. More complex systems require longer data sets to establish a statistical relationship between the response and predictors. However, caution should be taken not too extend to datasets over too long a window, as the model can be come difficult to interpret [50].

Table 7.3 displays some of the coefficients derived from the GLM with both the variable

Variable	Definition	Parameter Value	Residual Error	PF
G24	Gen Bus 144	0.0811	0.0143	1.000
G14	Gen Bus 52	0.0342	8.2e-03	0.4234
G15	Gen Bus 56	0.0361	7.7e-03	0.2177
G5	Gen Bus 16	0.0089	3.4e-03	0.2513
G8	Gen Bus 27	0.0062	1.3e-03	0.0453
G9	Gen Bus 29	0.0098	1.5e-03	0.0512
G21	Gen Bus 138	0.0107	2.3e-03	0.0512
G22	Gen Bus 139	0.0085	8e-04	0.0617
G5(MV)	Gen Bus 16	-0.0105	1.7e-03	0.1513

Table 7.3: Parameters and Residual Error for Selected Variables from 256 point optimal scale GLM with Reactive Power Components

selection applied and the reactive power variables added to the model. The inclusion of the reactive power variables has served to alter the β coefficients as expected but the important predictors have maintained their prominence. It is also evident that the reported predictors agree quite well with the participation factors displayed later in figure 7.5. The WGLM dictates that GB43, GB50 and GB139 are the most significant variables as they have the largest GLM regression coefficients.

An interesting point in table 7.3 is the inclusion of the reactive power variable 30 which

corresponds to GB16 MVAr output. Again, the reactive power variable has a small regression coefficient due to the fact that reactive power is normally not as well as correlated as active power variables. However, G5(GB16) MW has been reported as well which may indicate that G5(GB16) MVAr is an important variable in the model just like G18(GB104) from the Icelandic 0.21Hz case.

7.2.5 WGLM Logic Regression Model Fit

In this section, results are presented from the logic regression based on the five leaf models derived from the cross-validation which determined an optimum five leaf model for the 0.62Hz mode. The results are then tested against an eigenvalue analysis and the WGLM as well as compared to the normalized decay time constants plotted over a suitable time window.
The results are derived from the Icelandic 0.62Hz mode which has already been subjected to variable selection and pre-processing which leaves a total of 25 generators whose active power outputs are used in this study. The reactive power variables included in the model are also subject to thresholding and variable selection before they were implemented into the model.

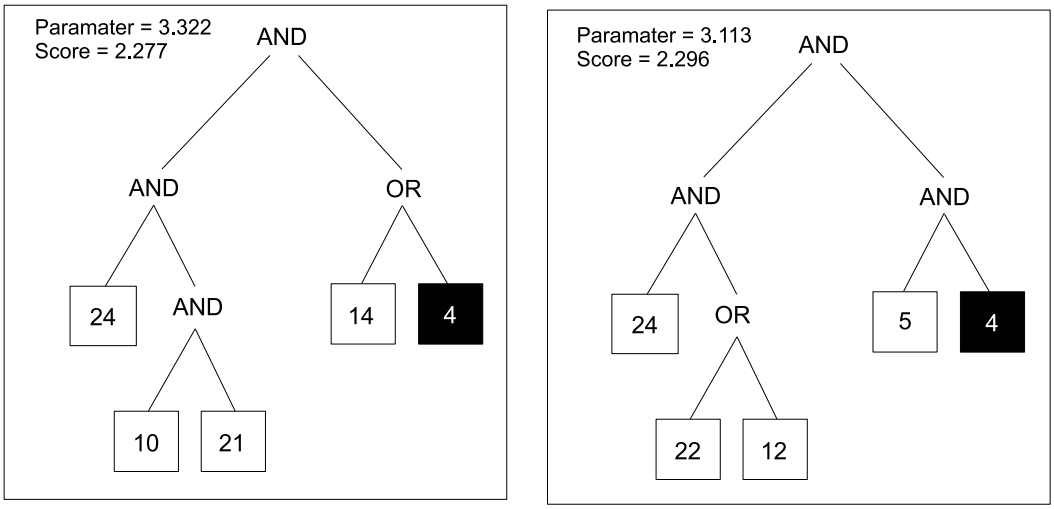
This model is based on the optimal scale combination detailed in section 7.2.4 with a 256 data length. This optimal scale GLM uses resolution levels 3,6 and 7 as these levels had the most significant wavelet coefficients when the signal was deconstructed. This optimal scale combination was shown to produce the best fitting GLM's in section 7.2.4 and as a result, were used in this study to determine the best logic models.

Table 7.4 shows the comparison of the best scoring model scores for the 0.62Hz logic regression using the various optimal scale combinations that were tested for the GLM's. It can be seen that model 6 which consists of the resolution levels 3,6 and 7 performed best as these were the resolution levels that also produced the best fitting WGLM's. This result differs from the 0.21Hz mode from chapter 6 section 6.4.5 where resolution levels 3,4,6 and 7 contained the majority of the dynamics.

Figure 7.4 shows a comparison of two five leaf models derived from the Icelandic

MCMC Run		
Model	Levels	Score
1	1,2,3,4,5,6,7	3.713
2	2,3,4,5,6,7	3.634
3	4,5,6,7	3.772
4	4,6,7	2.512
5	3,4,6,7	2.411
6	3,6,7	2.277

Table 7.4: Comparisons of model scores for different combinations of 256 length resolution levels



(a) Five leaf logic tree output

(b) Five leaf logic tree output

Figure 7.4: Comparison of five leaf output from the logistic regression showing some of the variation in the model. The numbers correspond to generator numbers

0.62Hz logic regression model. It can be seen how the outputs vary quite a bit with a total number of eight variables being represented between the models. The MCMC step allows all the variables in each model to be tabulated and compared against each other to determine how many times they appear in the same model. This in turn allows us to establish the most common interactions produced within a model as well as removing some of the redundant variables and interactions that also take place.

Table 7.5 shows the comparison of the number of observations of joint variables in the MCMC models. The single column shows a common set of single variables in the

MCMC Run		
Single	Double	Triple
24	14,5,15,21,22,23	15,8,21,19,22
14	24,8,19,15,21,5	15,8,21,19,22
15	4,21,24,8,14,19	24,21,8,14,10
21	15,5,22,24,4,14	4,22,14,24,5
19	14,24,15,19,8,18	5,14,5,15,19
22	15,5,24,14,4,22	4,5,24,14,19

Table 7.5: MCMC Comparisons from the most common variables derived from the five leaf wavelet transformed models

MCMC step with generator variable 24 (GB144) being the most common. These single variables have been selected as they form interesting interactions across the system. Other, more common single variables have been excluded to reduce the table size. The double column displays the most common variables that appear in the models alongside the single variable in column 1.

The MCMC single variables seem to be in good agreement with both the logic regression results shown in figure 7.4 as well as the results from the wavelet transformed GLM in section 7.2.4 with G14(GB52) and G24(GB144) being reported as significant. A large number of the common single variables also appear jointly in the model with other common single variables as evidenced by the double column.
For example, generator variable 14 (GB52) also appears quite regularly with variable G24 (GB144) which is a also common single variable. Other pairings include variable G14 which is jointly observed with variable G8 (GB27), however, variable 8 is not shown in table 7.5 as only six variables and their interactions were displayed.

However, this is the benefit of using the MCMC step which is used to highlight these coincidences across a number of models. This allows the less common variables to be included in the model by virtue of association with common variables.
The triple values also offer some further information on the frequency of groups of variables in the model. In the case of the significant variables G14 and G24 (GB52 and GB144), it can be seen that variable 5 (GB16), a common single variable, is observed quite often with this pair. In this case, it may be useful to look at the effect this

combination of generators has on damping.

Figure 7.5 displays the average participation factors for the Icelandic 0.62Hz mode

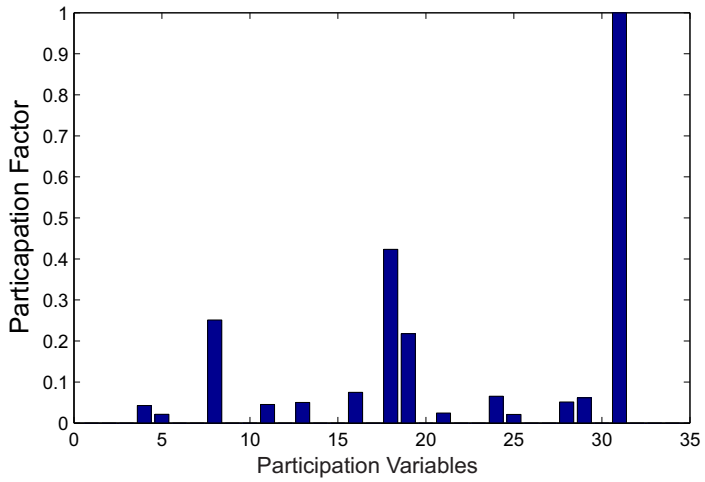


Figure 7.5: Icelandic 0.62Hz Mode Participation Factors derived from rotor angle states

under consideration. In section 2.7 participation factors were introduced as the product of the left and right eigenvectors for a given state (rotor angle, rotor speed etc.) and mode respectively. By calculating the participation factors for the generator rotor angles we can determine which generators can be used to control a mode. Generator active power is proportional to the sinusoid of generator rotor angle, so by controlling the output power of a participating generator we can control the mode.

The results in figure 7.5 have been normalized with the largest value attributed to participation variable 31 which corresponds to G21 (GB139) in the reduced model after variable selection. This is explained in section 6.4.6 in reference to figure 6.4, where the participation variables are referenced to generator numbers. It must be noted that the 34 variables in the plot represent the 34 generators in the model prior to variable selection. With the selected variables removed, 24 generators are left corresponding to G1-G24 which are used to develop the statistical model.

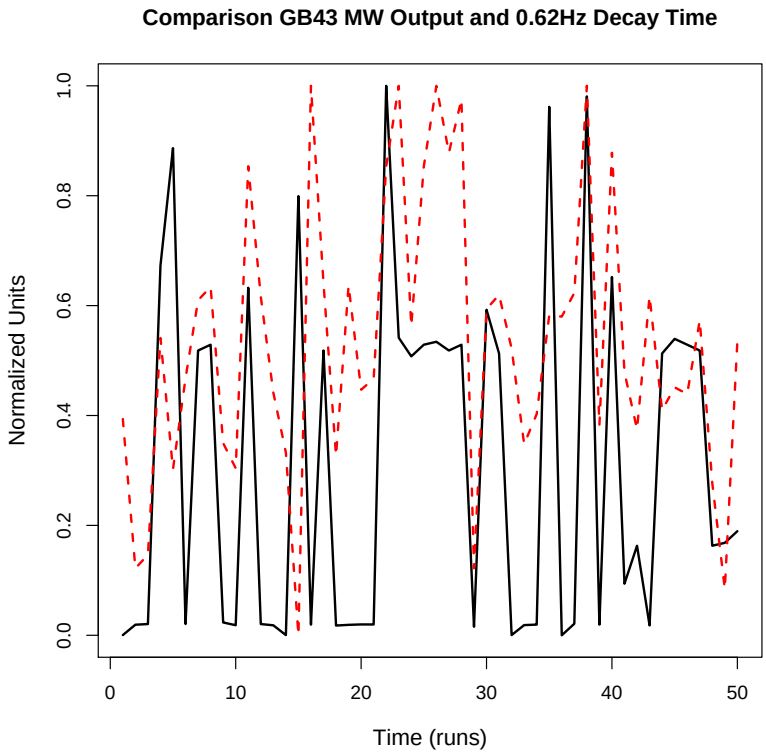
As expected, there is some agreement between the participation factors and the logic

regression step from figure 7.4 and table 7.5. These results are similar to those in chapter 6 section 6.4.6 which compared the 0.21Hz logic regression results to the participation factors and WGLM results and found reasonable agreement. In this case, variables G8,G14 and G24 (GB16, GB52 and GB144) along with variable G15 (GB56) have a higher than average participation factor.

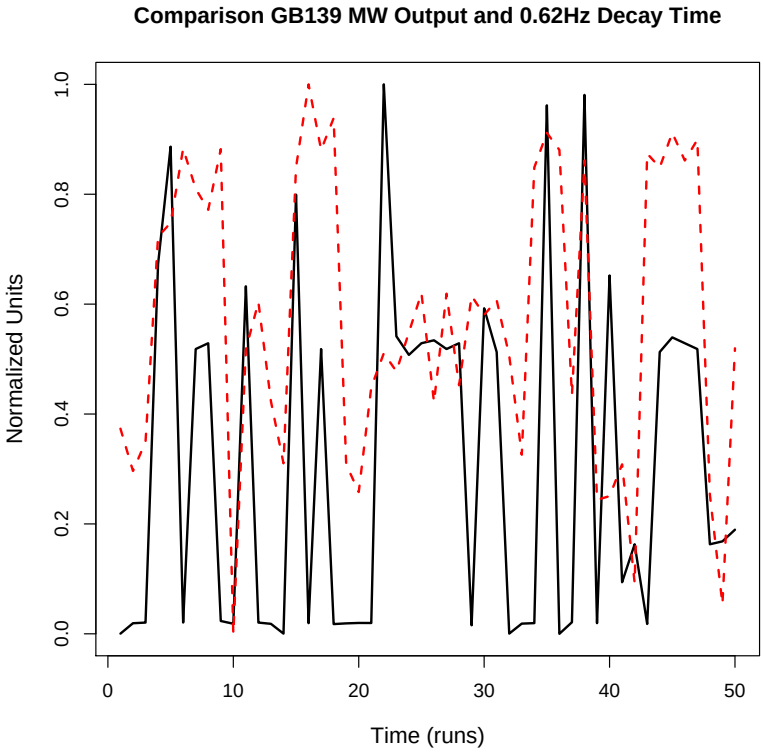
In figure 7.6 the subplots show the superposition of the 0.62Hz mode decay time constants against the generator outputs for bus G14(GB52) and G22(GB139) which were the two most significant machines reported in the logic regression and the WGLM. The most significant machine outputs (G14(GB52) and G24(GB144)) seem to correlate with all the events as either or both of them are high when the decay time constant is poor.

By using a simple linear regression, single effects may be considered i.e. a single predictor having an effect on damping. In the logic regression case, we can see that at least two variables have an effect on mode damping. However, the logic trees in figure 7.4 also suggest that variable G20 (GB126 and GB127 as it is an aggregated generator from generators connected to two separate buses) or indeed variable G22 (GB139) may have an effect on the mode damping. As a result, it is worth looking at the effects a reduction in these generator outputs might have on the mode damping.

Figure 7.7 shows the selected decay time constant profiles when G22(GB139) is replaced simultaneously with a constant power source reduced to 25% of its nominal value. This plot shows how dependent the 0.62Hz mode is on the machine connected to G22(GB139). A large number of the mini-events have now been removed from the profile with only three of the original decay time constant spikes remaining. It is clear from the WGLM and the logic regression that GB139 was an important machine in relation to the 0.62Hz mode.
However, the remaining spikes in the profile are more interesting as in these instances the decay time constant is not solely dependent on G22(GB139) but a combination of machines. It is this combination of machines that will be extracted using logic regres-



(a) Bus B43 Comparison



(b) G22(GB139) Comparison

Figure 7.6: Plots showing significant generator outputs (red) against 0.62Hz mode decay time constant

sion and tested in the next part of this section.

Figure 7.8 shows the selected decay time constant profiles when G22(GB139), G20(GB126)

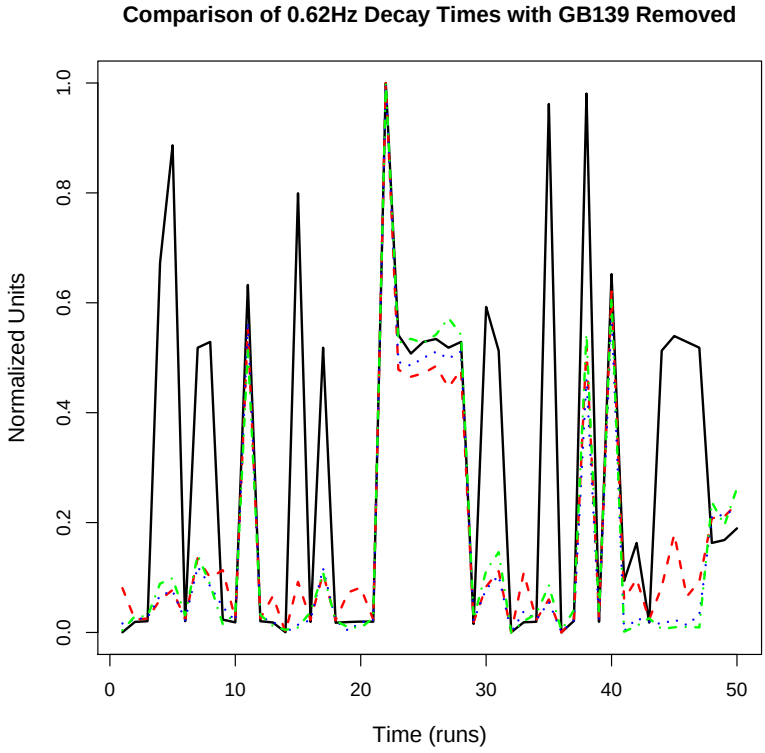


Figure 7.7: Comparison of selected decay time constant plots when G22(GB139) is modulated

and G20(GB127) (split over two buses) are replaced simultaneously with a constant power source with the output reduced to a fraction of their nominal values. This plot is very revealing as the effects of simultaneously reducing the three generators can be seen in the fact that all of the original mini-events have completely disappeared from the profile. In comparison to figure 7.7 it can also be seen that two new mini-events have been created.

This can be explained as follows. In the logic trees of figure 7.4 and the MCMC table in table 7.5 it was shown how generator variables 21 and 22 appeared in conjunction with generator variable 24 in a number of the logic models. In actual fact, variable G22 wasn't as common as variable G21 but was deemed significant as the models it

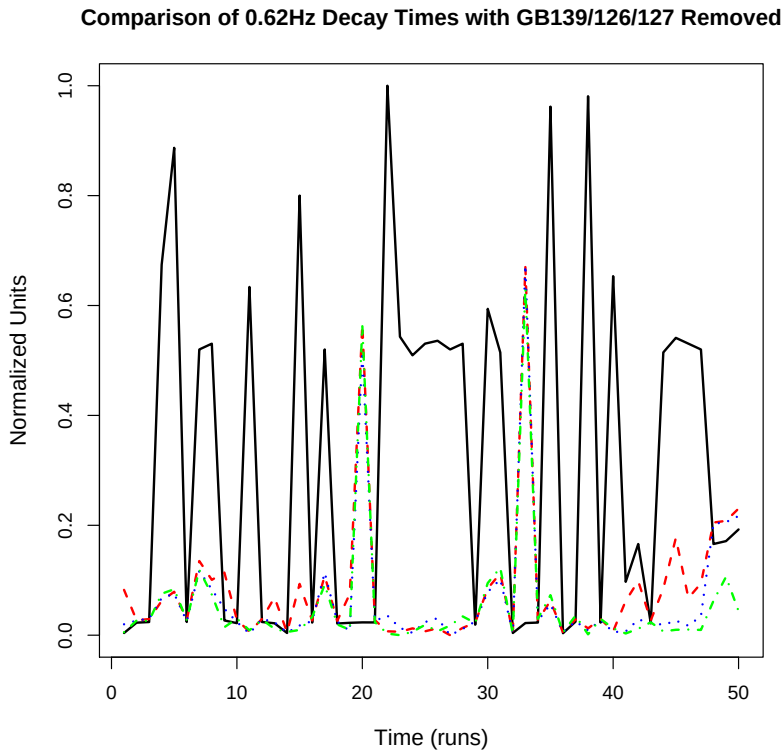


Figure 7.8: Comparison of selected decay time constant plots when G20 (aggregated GB126/GB127) and G22 (GB139) is modulated

appeared in contained significant variables for the most part.
Since G20 (GB126 and GB127) did appear together with high ranked variables, it was clear that they must be linked as they are located in the same plant. As a result, it was discovered that these two machines formed an AND pair which when deactivated with G22(GB139) caused the original decay time constant record to be reduced below a 3-4 second base value in comparison to a 22 second maximum, measured previously.

This suggests that the interaction of G20 and G22 (GB126, GB127 and GB139) have a strong effect on mode damping as suggested by the MCMC table and logic trees. Since these poorly damped events are removed when both machines are reduced, this means that both these machines are related to the damping and can be used to provide a more robust control of the mode dynamics.
However, this is one of the disadvantages of the incomplete methodology. The AND pair of G20 (GB126 and GB127) was not explicitly stated in the logic trees or the MCMC table. It was through the testing of a number of variables and observing the interactions that this AND pair was discovered, not through direct observation. Although this could be incorporated by automatically testing the combinations of variables from the MCMC tables, some engineering judgment is required to extract some of the key interactions.

The creation of new "mini-events" in figure 7.8 may be due to the action of G13(GB43) or a combination of other machines in the network that have not been modulated. Since the model is setup with random injections and intentional large scale injections, as explained earlier in section 4.6.2, the modulation of a number of generators may cause other events to occur along the testing period. In this case, the mini-events may be dependent on G13(GB43) interacting with other significant generators such as G5(GB16) or G15(GB53). In this section, we have only seen one interaction that affects the 0.62Hz mode damping but it is highly likely that there are others.

It has been shown in this case that a number of variables are linked to damping across the window shown. G20 and G22 (GB126, GB127 and GB139) were shown to have

a significant effect as the decay time constant is substantially reduced only when all three machines are reduced. Although the interaction is not explicitly stated in the methodology it can be deduced using some logic and used to control mode damping. In addition, this may not be the only interaction that has an effect on mode damping. In chapter 6, two separate interactions were shown to have an effect on the 0.21Hz mode. This again shows the effectiveness of the method where a number of different machines can be used to control mode damping.

This is highlighted by the fact that some new mini-events occurred as shown in figure 7.8. These mini-events were created after the modulation of the G20 and G22 (GB139, GB126 and GB127) combination. Although efforts were made in the modeling process to keep the subsequent model runs as constant as possible, by modulating the same large power injections the complexity of the network is such that other generators and combinations thereof may trigger new events. The presence of these events proves that other combinations of generators interact to produce dynamics at different instances along the sample window.

Figure 7.9 shows the decay time constant plots from a number of loadflow runs with generators G13, G15 and G22 at generator buses 43, 53 and 139 modelled as constant power sources so that they output 21% (GB139 set at 25%) of their nominal output irrespective of the dynamic conditions of the system. These generators represent the primary interaction derived from the wavelet transformed logic model. In comparison to figure 7.8 it can be seen that the primary interaction is the most successful in that the long decay time constants are totally diminished.
However, in order to obtain a more robust control of damping it is worth testing a number of interactions to see if system stability can be improved. In this case, the primary interaction sufficed as a control but this is not always the case. It may be worth amalgamating two or more interactions using the MCMC step to determine the best course to alleviate poor damping.

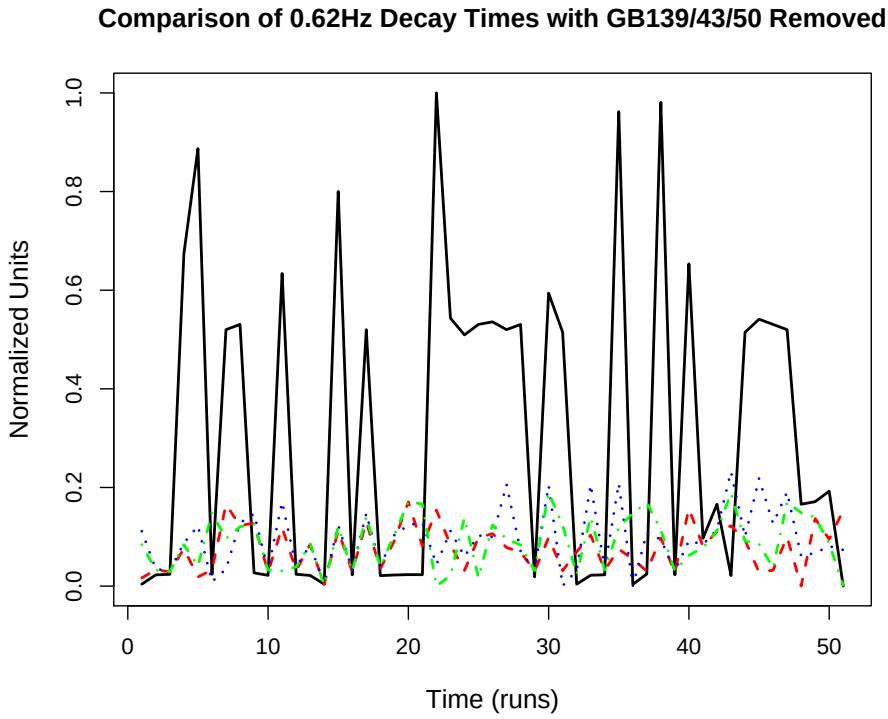


Figure 7.9: Comparison of selected decay time constant plots when G22 and G13 (GB139/GB43/GB50) when modulated

7.3 Development of Raw Logic Model

The development of the raw logic model takes place in a similar fashion to the wavelet transformed model. The raw model is used in this section, as previous studies indicated the existence of mutually exclusive significant predictors that appear in both the wavelet transformed and raw models.

The variable selection methods outlined in chapter 5 section 5.3.2 will be employed throughout this section so as to improve the logic model prediction. Once the raw logic model has been developed, the normalized plots of the reported predictors will be plotted against the normalized decay time constants to see if the results can be validated.

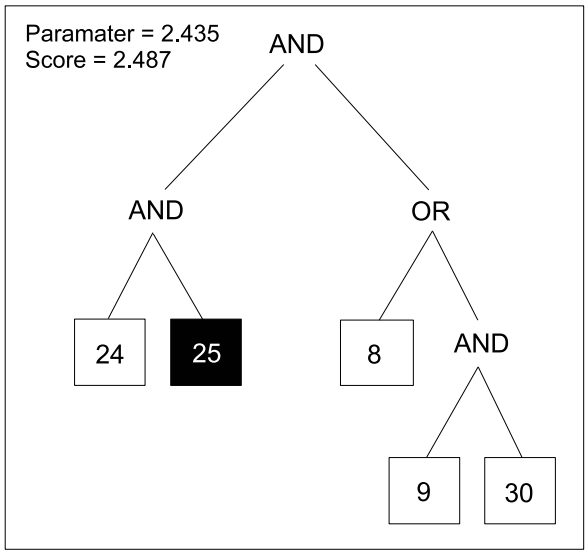
7.3.1 Development of Logic Expression for Raw Icelandic Model

The raw data is more suited to a logistic logic model as the response/predictor relationship is non-linear. As a result, the entire raw dataset (response and predictors) has been transformed into binary values and used in the logic regression.
Variable selection has already been applied, with the remaining variables being used in this study. Reactive power has also been included in the logic model as it has been shown to improve the model fit.
The key point in this section is the issue of highlighting the difference between the raw and wavelet transformed data outputs from the respective logic regressions. To follow on from this, both sets of results need to be reconciled and aggregated due to the fact that both sets contain valid observations that may be used to control mode damping.

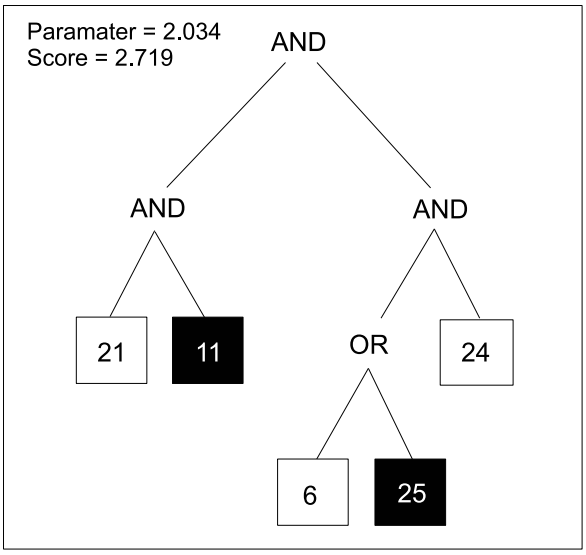
Figure 7.10 shows a comparison of logic trees derived from the initial runs for the 0.62Hz raw data. Subfigures 7.10(a) and 7.10(b) show two of the logic trees produced from the same thresholded data. Since logic regression relies on a number of algorithms to search through the state space there is always some variability in the logic functions that are produced.

Familiar generator variables that were reported in subfigures 7.10(a) and 7.10(b) include variables G21 and G24 which correspond to generators connected to buses 126 and 139. These were also highlighted in the WT model in the previous section.
For these logic trees, it can again be seen that there is partial agreement between the wavelet transformed logic models, the raw logic models and the wavelet transformed GLM presented in sections 7.2.4 and 7.2.5.

However, the raw logic model presents different interactions in comparison to the wavelet transformed models. As well as this, some of the significant variables from the wavelet models are less common in the raw models and vice-versa. This subtle difference may have interesting consequences for the source location methodology as more variables are made available for testing and validation. By aggregating and com-



(a) Five leaf logic tree output



(b) Five leaf logic tree output

Figure 7.10: Comparison of outputs from raw logic regression step

bining both the raw and wavelet transformed results it is hoped that a more robust description of the system dynamics can be presented.

Table 7.6 shows the comparison of the number of observations of joint variables in the

MCMC Run		
Single	Double	Triple
24	25,8,21,11,30*	8,9,21,30*,11
25	24,9,8,6,21	8,9,21,30*,11
8	9,24,21,25,30*,	24,25,6,11,21
21	8,24,25,21,11	9,25,24,6,11
9	8,25,24,11,21,30*	24,25,6,11,21
11	24,9,25,21,8	25,9,8,6,21

Table 7.6: MCMC Comparisons from the most common variables derived from the five leaf raw models

MCMC raw models. The single column shows six of the most common variables in the MCMC step with variable G24 (GB144) being the most common. The MCMC single raw variables seem to be in good agreement with both the wavelet transformed logic regression results shown in table 7.5 as well as the results from the wavelet transformed GLM shown in table 7.3 comparison.

This agreement is strengthened by the fact that generator buses G8(GB27) and G25(GB146) are reported as significant. A large number of the common variables also appear jointly in the model with other common variables as evidenced by the double column. For example, variable G8 (GB27) also appears quite regularly with variables G24 and G25 which are also common single variables. In addition, it can be seen that variable G8 is also jointly observed with variables G9 (GB29) and G30 (GB16-MVAr) even though variables G9 and G30 are not all that common in comparison to the top ranked variables.

Variables G21 and G22 were common variables in the wavelet transformed logic models that were validated via the MCMC step. Apart from these variables, the raw and wavelet transformed models are quite different and complementary, even moreso than the 0.21Hz case reported in chapter 6 section 7.2.5. We have already seen how some

of the peripheral variables in the wavelet transformed logic model, such as GB126 and GB127, have a significant effect on the mode damping. It may be wise to test similar peripheral variables in the raw model to see if similar interactions can be used to control mode damping.

Figure 7.11 displays the average participation factors for the Icelandic 0.62Hz mode

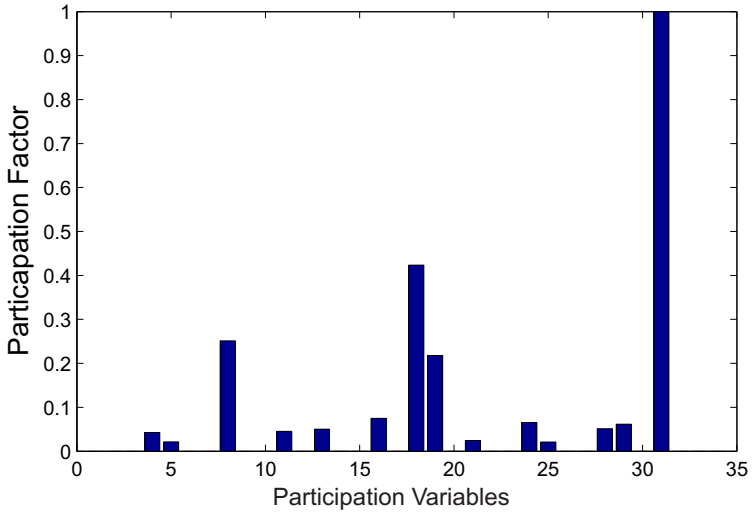


Figure 7.11: Icelandic 0.21Hz Mode Participation Factors derived from rotor angle states

under consideration. The results have been normalized with the slack bus removed as explained in chapter 6 section 6.4.6. The largest value of participation factor has been attributed to variable 31 (G22) which corresponds to GB139.

As expected, there is some agreement between the participation factors and the raw logic regression step with G22(GB139) having the joint honour of being the most prominent logic regression variable as well as having the largest participation factor. Other machines reported via the logic regression have quite small participation factors, including G8(GB27) and G9(GB29) which correspond to participation variables 11 and 13 (change of indexing caused by variable selection).

Participation variables 18 and 19 from figure 7.11 correspond to generators G13(GB43) and G15(GB53). G13 and GB15 are ranked 2nd and 4th in the participation factor

column of table 7.7. However, they are not reported by the logic regression step as they do not form interactions with other variables to cause poor damping. One of the goals of this section is to determine the effect these lower ranked machines have on the damping in conjunction with the more prominent generators. This must be studied as there is some disrepancy between the logic regression step and the participation factors, the source of which may lie in the interaction effects of the generators.

Table 7.7 shows the comparison between the aggregated logic regression variables and

MCMC Run		
Simulation	LogicReg	PF's
Gen. +/- Values	22,25,8,20,9,B33	22,B43,3,13,16,20
Gen. Bus +/- Values	139,145*,27,126,29,B33	139,B43,16,43,99,126
Gen. - Values	22,8,20,9,10,3(MV)	22,B43,3,13,16,20
Gen. Bus- Values	139,27,126,29,33,16(MV)	139,B43,16,43,99,126

Table 7.7: Significant variable comparison between raw logic regression and participation factor results

the participation factors for the 0.62Hz mode. The first two rows of results contains negative Boolean variables for generators and generator buses i.e. GB144 in a NOT value, whereas rows 3 and 4 contains the ranked logic variables with the NOT values removed. It can be seen that there is some agreement between the logic model and the PF's with the presence of G22(GB139), G20(GB126) and G3(GB16) (MW in participation factors and MVAr in the logic regression). This result suggests that participation factors require supplemental information as they cannot fully account for many-way interactions in the power system. This was seen in the fact that G8(GB27) and G9(GB29) (which are important interactive variables) have negligible participation factors.

Figure 7.12 shows the selected decay time constant profiles when both GB139 as well as G8(GB27) and G9(GB29) are replaced simultaneously with a low constant power source independent of the system dynamic conditions. It can be seen how in subfigure 7.10(a) variables 8 and 9 (GB27 and GB29) are connected via an OR statement in the logic tree. However, owing to some irregularity in the Boolean operators, where AND

and OR operators linking G8 to G9 (GB27 to GB29), frequently interchanged it was decided to see the effect that modulating variables 8 and 9 simultaneously would have on the 0.62Hz mode i.e. modulating variables G8 and G9 as an ANDED pair.

Again, this plot is quite revealing as a number of "mini-events" have been dimin-

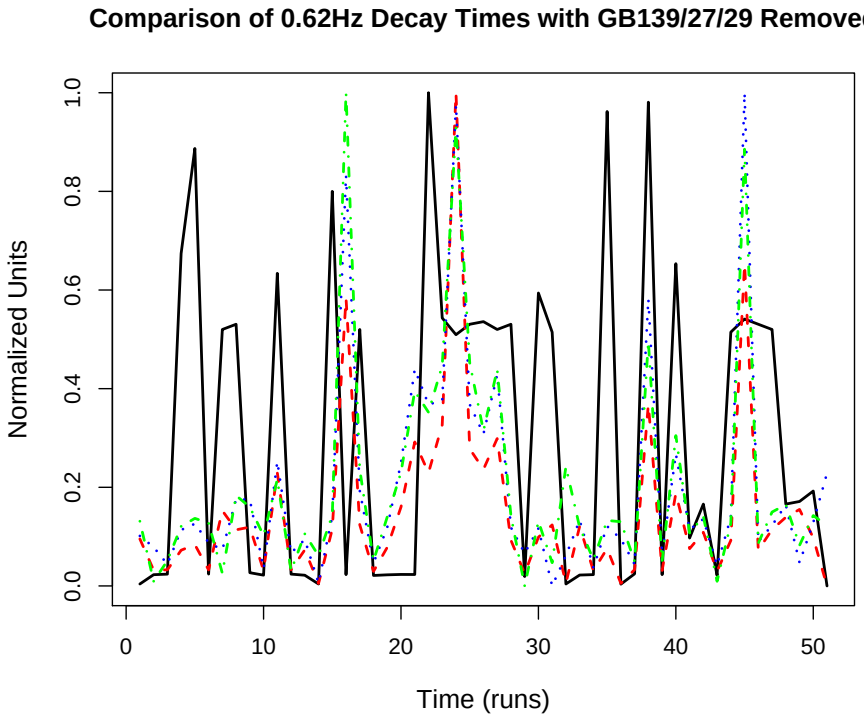


Figure 7.12: Comparison of selected decay time constant plots when G8(GB27), G9(GB29) and G22(GB139) are modulated

ished in comparison to figure 7.7. An example of this is the mini-event between 20 and 30 seconds which has been partially reduced. This is quite typical of the logic regression methodology in that the modulation of the reported machines produces partial degradation in mode decay time constants. It can be seen here that this mini-event has been partially degraded due to the modulation of all three generators (G8(GB27),G9(GB29),G22(GB139)) but some of the large decay time constants still remain.

This result suggests that the interacting generators reported by the raw logic regression do indeed have an effect on the 0.62Hz mode damping. However, the logic tree

in subfigure 7.10(a) suggests an alternation between AND/OR operators in defining the G8/G9 (GB27/GB29) action whereas it has been shown that when both of these machines are reduced, the mode is partially stabilized. This is where the MCMC step in table 7.6 is useful as it highlights common combinations of variables that can be tested to complement the straight logic tree results.

It is obvious that both the raw and wavelet transformed logic regression methodologies are linked as they both have a number of variables in common. In this case, the wavelet transformed model seems to perform better, as the dominant interaction can be used for a more robust control of damping. This is due to the fact that the decay time constants are significantly reduced in comparison to the raw model. This can be confirmed by comparing figure 7.9 and figure 7.12.
There is a considerable amount of overlap between the raw and wavelet transformed logic regression models. Seeing as the interaction effects evolve over time, it can be difficult for a given logic model to encompass all sensitivities taking place in the system.

7.4 Case Study: Australian Power System

7.4.1 Australian 0.48Hz Interarea Mode

The 0.48Hz mode measured at the Strathmore feeder in the Australian system between 27/12/07 - 03/01/08 exhibited some excessive decay time constants with a maximum of 75 seconds being measured during this period. Active and reactive power variables as well as bus voltage variables were recorded during this period in order to determine the causes of poor damping.
In this case the generator data was split up into five different regions namely Queensland, Victoria, New South Wales, Southern Australia and Inter-Area tie data. In this work, each region is taken separately to allow a individual statistical models to be developed as the entire system consists of over 200 generators. If these were included in a single model, problems with dimensionality would cause distortion of the prediction and the important relationships between the predictor and response variables.

Initially, the data from each region was subject to the variable selection procedures. This data was then used to construct logic models from which the important interactions were extracted for use in controlling damping. Important interactions from each region along with the model scores are compared to determine the best variables to use to control damping.
Again, optimal scales were used in the logic models to remove the redundant frequency levels which are derived from the wavelet deconstruction of the decay time constant values. The raw logic models were then compared to the wavelet models to find any mutually exclusive interactions that may be used to control damping.

Finally, the significant variables and interactions reported for various data lengths were plotted against the 0.48Hz mode damping for comparison to determine any correlation between them. Seeing as there is no way to test the relevant generators it is hoped that the validation process outlined in chapters 5 and 6 as well as at the start of this chapter, show that the method can be used to establish the important interactions that create the conditions for poor damping.

7.4.2 Australian 0.48Hz Results

In this section, results are presented from the logic regression based on the various sized models derived from each of the regions in the Australian system. Comparisons of the normalized decay time constants plotted against the normalized predictor values over a suitable time window are also presented to validate the results.

This model is based on the optimal scale combination with a data length of 2048 data points. This optimal scale logic model consisted of levels 6,7,9 and 10 as these levels had the most significant wavelet coefficients when the decay time constant signal was deconstructed. As a result, the corresponding frequency bands of the active and reactive power signals were also used to develop the statistical models. These levels corresponded to a total of 1728 out of a possible 2047 datapoints for a full scale model. This optimal scale combination should produce better fitting models in relation to the

full wavelet spectrum as the redundant frequency bands result in a distorted model.

Table 7.8 shows the comparison of the best scoring model scores for each of the five

MCMC Run		
Region	Levels	Score
Queensland	6,7,9,10	3.032
Victoria	6,7,9,10	2.342
NSW	6,7,9,10	3.144
South Aus	6,7,9,10	3.712
InterArea	6,7,9,10	3.580

Table 7.8: Comparisons of model scores for different combinations of 256 length resolution levels

regions in the Australian system. These models are based on the resolution levels 6,7,9 and 10 from before. From this table it can be deduced that the best fitting model is obtained when the Victoria data is used, with the next best fitting model being the Queensland model. As a result, the Victoria model will be used first to determine the cause of the 0.48Hz damping issues.

Figures 7.13 and 7.14 show a comparison of two of the models derived from the Icelandic 0.62Hz Victoria data. It can be seen that there is quite a bit of continuity between both the three and five leaf models as variables 3, 14 and 19 appear in both models.

Table 7.9 shows the comparison of the number of observations of joint variables in the

MCMC Run		
Single	Double	Triple
19	3,14,13,1,11,7	14,13,10,11,1
14	19,3,13,10,1,5	3,13,1,11,10
3	19,14,11,10,13,12(MV)	14,11,13,10,6
1	11,19,14,7,10,3	19,3,14,7,10
11	1,7,19,22,14,10	19,3,14,7,10
13	7,5,14,11,19,11(MV)	5,19,14,11(MV),24

Table 7.9: MCMC Comparisons from the most common variables derived from the models

MCMC models derived from the five leaf setup. The single column shows a selection

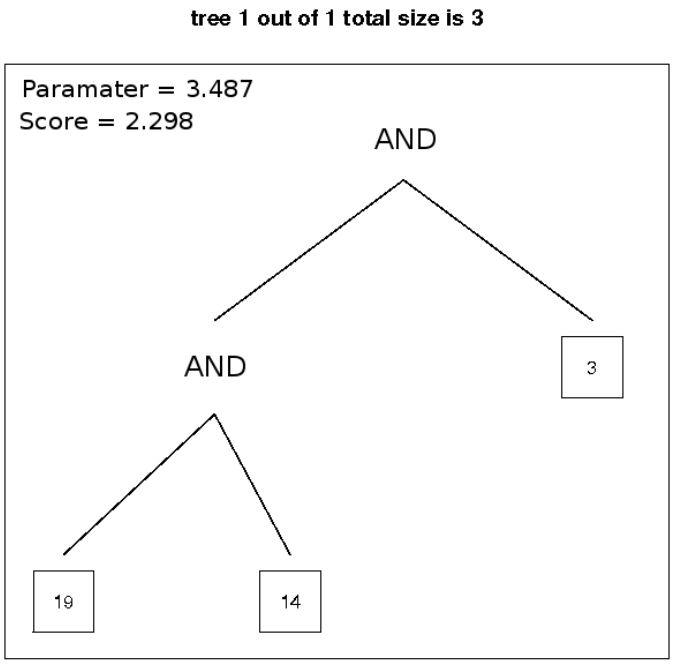


Figure 7.13: Three leaf output from the logistic regression showing some of the variation in the model

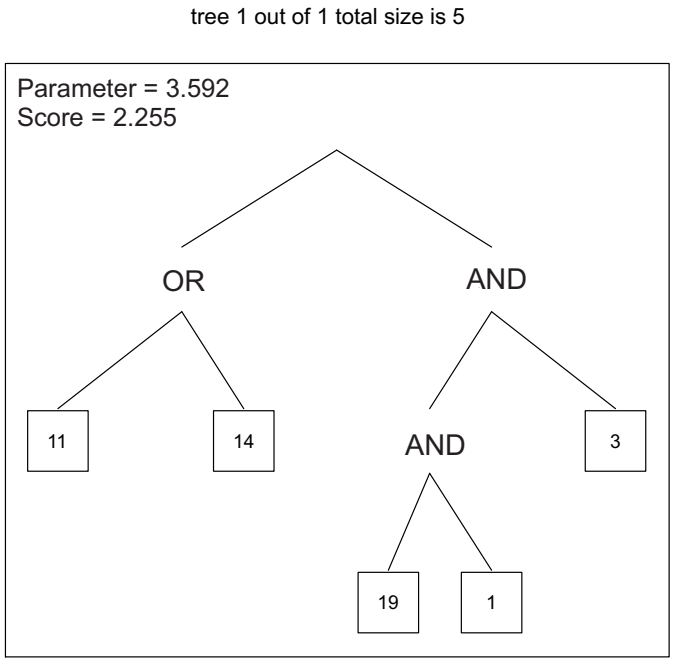


Figure 7.14: Five leaf output from the logistic regression showing some of the variation in the model

of six common variables in the MCMC step with variable 19 (generator 22) being the most common.
Both figures 7.13 and 7.14 also display the operators that link the important system variables to form the interactions. The two interactions shown are typical of the three and five leaf models where variables 3,14 and 19 are always connected via AND operators in both models. They are usually connected via AND operators in the five-leaf model where either variable 11 or 14 can be AND'd with variable 1, 3 or 19. These variables, as well as others will be plotted later to see if they can be validated.

The triple column in table 7.9 contains variables that appear with the two most common variables reported in the single and double columns. These can be tested along with the single variables to determine some of the peripheral variables that interact to cause poor damping. It must be noted that these peripheral variables are not always truly significant and may not be classified as common single variables. However, from previous examples it was seen how these variables can be extracted from the MCMC tables and used to control mode damping quite effectively.

The MCMC single variables seem to be in good agreement with both the logic regression results shown in figures 7.13 and 7.14, which is to be expected as they are derived from the same logic regression models. A large number of the common single variables also appear jointly in the model with other common single variables as evidenced by the double column.
The MCMC step is useful in highlighting less common variables to be included in the model by virtue of association with more common variables. This is illustrated with variables 7 and 13 (G7 and G14) as they are not high ranked single variables, but are present in quite a few models containing significant variables.

7.4.3 Validation of Results

In this section, the reliability of the results from the previous section will be tested by comparing the generator outputs and the decay time constant profiles of the 0.48Hz

mode.

Referring back to figures 7.13 and 7.14, the score and parameter values have been shown for both the three and five leaf models. Both values are comparable to the values for the Icelandic examples in chapters 6 and 7 which would indicate quite a well fitted model. It is evident that the parameter value has increased and the score value decreased for the five leaf model in comparison to the three leaf model. This increased parameter value and decreased score value is due to the fact that by using more significant variables in a tree, the more of the response variable variance can be explained by a given interaction and thus the higher the parameter value. However, an excessively large logic tree may serve to reduce the model accuracy as all the reported variables may not be statistically significant.

The Victoria data allows a well fitted model to be developed. As a result, there is little variation in the logic tree outputs in the three and five leaf cases i.e. the same trees are produced after each run. However, the MCMC step still produces variant models which allows the MCMC step to determine which variables are jointly observed across the MCMC model space.

In this section the three most prominent variables will be tested to determine if they are significant or not. This group consists of variables 3,14 and 19 (G3,G15,G22). In addition, variables 1,7,11 and 13 (G1,G7,G11 and G14) will also be tested as they have been reported as significant either directly or by association with significant variables. In figure 7.15 the subplots show the superposition of the 0.48Hz mode decay time constants against the generator outputs for G3, G15 and G22 as these were reported as the most significant variables from the logic model as well as from the MCMC step. It can be seen how the three generator outputs remain above the decay time constant plot for most of the window. In the case of G22, it seems to be very well correlated with the decay time constant as it almost seems to track the decay time constant signal across the window. Generators 3 and 15 have also been included as they are both in a "high" state during the poorly damped event. In addition, both generator outputs display some interesting correlations, in that the initial increase in generator 15 at approximately 100 minutes coincides with the increase in the 0.48Hz decay time constant.

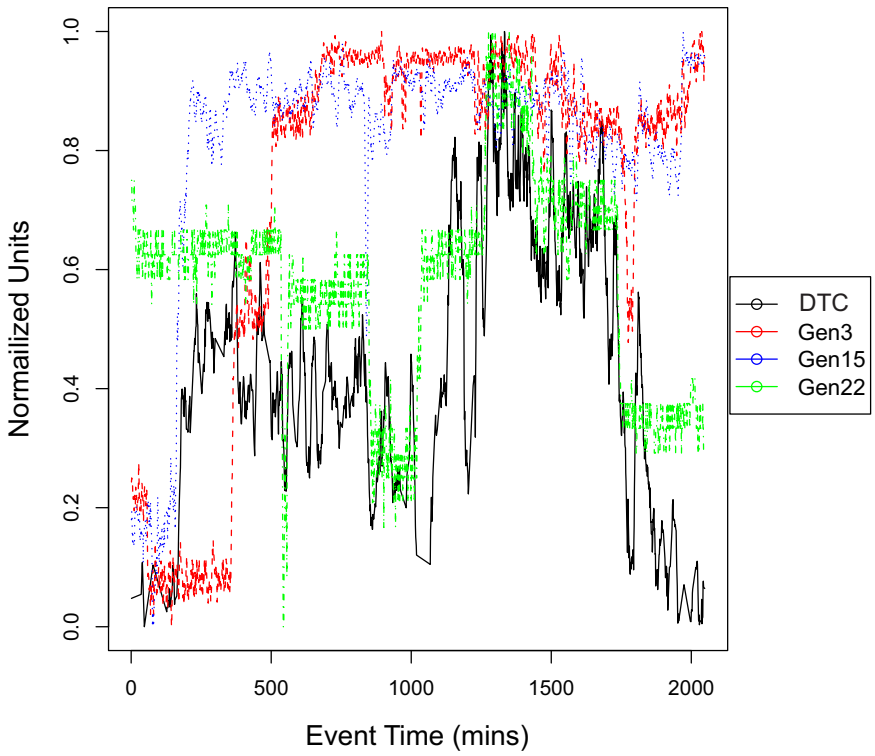


Figure 7.15: Australian 0.48Hz Mode Superimposed with Significant Predictors from Victoria State. DTC stands for decay time constant

Generator 3 also remains high throughout the window which may indicate that the G3 output coupled with the high outputs of generators 15 and 22 are what cause the poor damping.

The results suggest that generator 22 is the most significant machine of the three reported as it is well correlated with damping. However, since G15 seems to almost trigger the event it is thought that it may have an effect on the 0.48Hz mode dynamics i.e. it may be part of a generation pattern that causes G22 to be strongly linked to damping. Finally, G3 was reported due to its high output (590MW) for most of the window. It may be the case that this large output coupled with the actions of G15 and G22 cause the poorly damped events witnessed in the system.

In figure 7.16 the subplots show the superposition of the 0.48Hz mode decay time constants against the generator outputs for G1,G7,G11 and G14. Subfigure 7.16(a) displays the relationships between the 0.48Hz mode damping and the G1 and G7 outputs. The G1 output is similar to the G3 output in that it has a high output (580MW)

for most of the event. G7 is a more ambiguous result as there is not much correlation between this predictor and the response. However, the G7 output has a mean of 220MW so it is possible that the G7 output contributes to damping from around 400-800 minutes into the event, although it is not as prominent as the variables reported in figure 7.15.

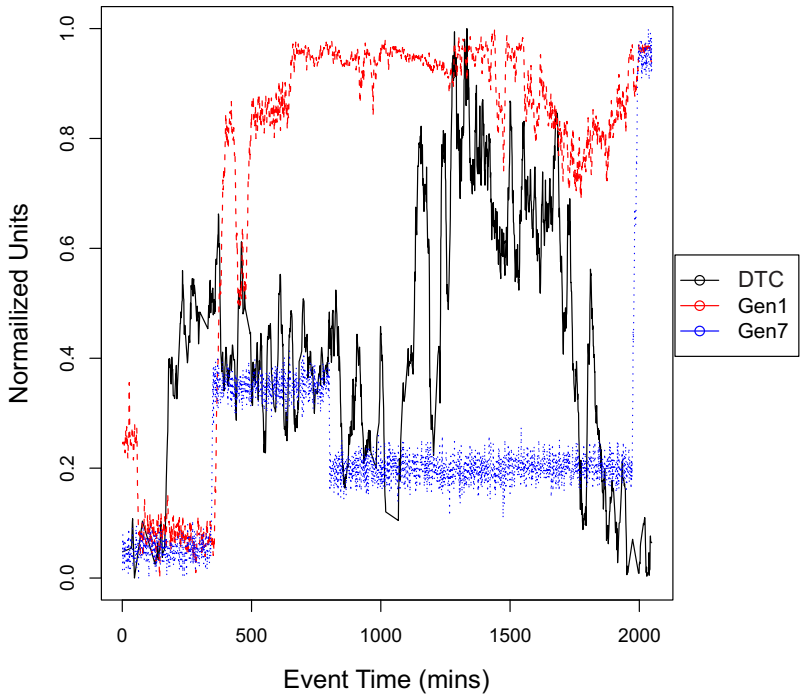
Subfigure 7.16(b) displays the relationships between the 0.48Hz mode damping and the G11 and G14 outputs. Again, the G11 output is similar to the G15 output in that it goes high at the same time the poorly damped event begins. The G14 output correlation with the decay time constant is ambiguous although its initial ramp up coincides with an increase in damping. G11 is similar in that it is sporadically correlated with the damping but is quite ambiguous. However, the G1, G7 and G11 interaction may be used as a secondary control along with the primary interaction reported in figure 7.13.

7.5 Development of Raw Logic Model

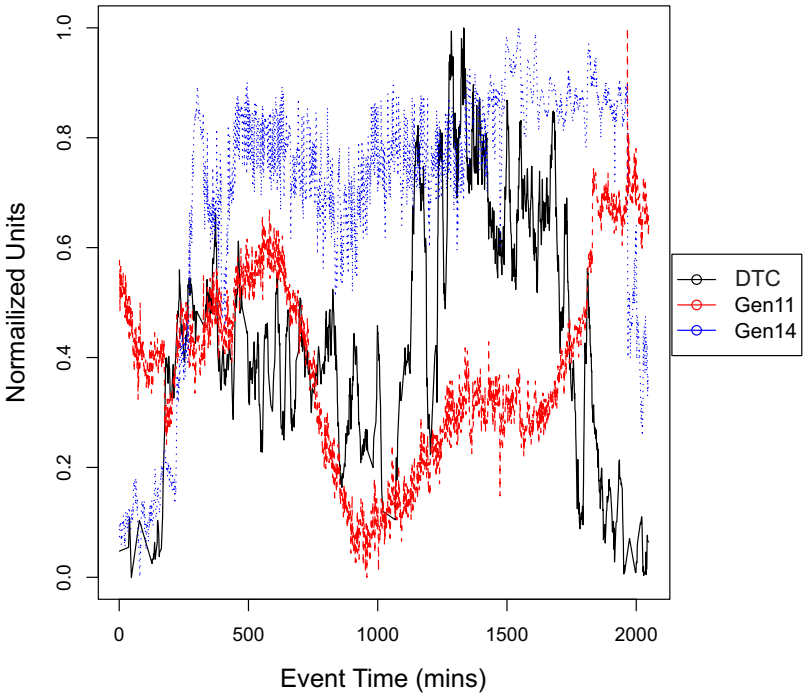
The development of the raw logic model takes place in a similar fashion to the wavelet transformed model with some slight differences. First of all, the raw model presents less of a challenge in terms thresholding as the variance of the data is less than that of the wavelet transformed data.
The variable selection methods outlined in chapter 5 will be employed throughout this section so as to improve the logic model prediction. Once the raw logic model has been developed, the results will be tested and validated by comparing the reported generators against the 0.48Hz decay time constants.

7.5.1 Development of Logic Expression for Raw Australian Model

In this subsection, the 0.48Hz Australian mode was investigated using the raw logic regression to determine the interactions leading to poorly damped events. The 2048 length dataset was used in this study as this setup produced the best results and allowed



(a) decay time constant Comparison with Gen1 and Gen7



(b) decay time constant Comparison with Gen11 and Gen14

Figure 7.16: 0.48Hz mode compared against significant predictors

a full poorly damped event to be included in the analysis. The raw data is more suited to a logistic logic model as the response/predictor relationship is non-linear. As a result, the entire raw dataset (response and predictors) are transformed into binary values and used in the logic regression.

Figures 7.17 and 7.18 show a comparison of two of the models derived from the Victoria

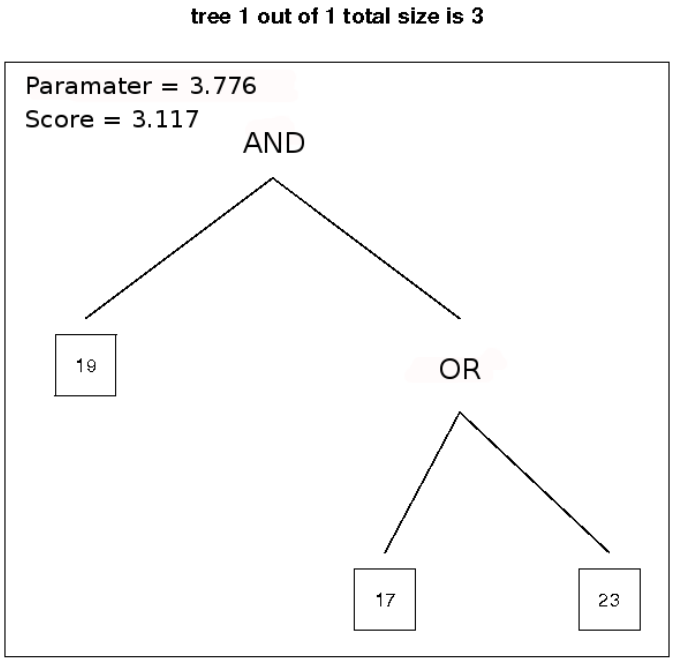


Figure 7.17: Comparison of outputs from raw logic regression step

0.48Hz data. It can be seen that there is quite a bit of continuity between both the three and five leaf models as variables 19 and 23 appear in both of the examples. Table 7.10 below displays the most common variables and variable groups derived from the five leaf models presented earlier. The most common variables reported in figure 7.17 are also reported in the MCMC table as variables 17,19 and 23 which correspond to generators 20,22 and 34.

Figures 7.17 and 7.18 display the operators that link the important system variables that form the interactions that affect mode damping. The two interactions shown are readily reproducible with variable 19 always being connected via AND operators to various other variables that may be connected via OR or AND operators at the second

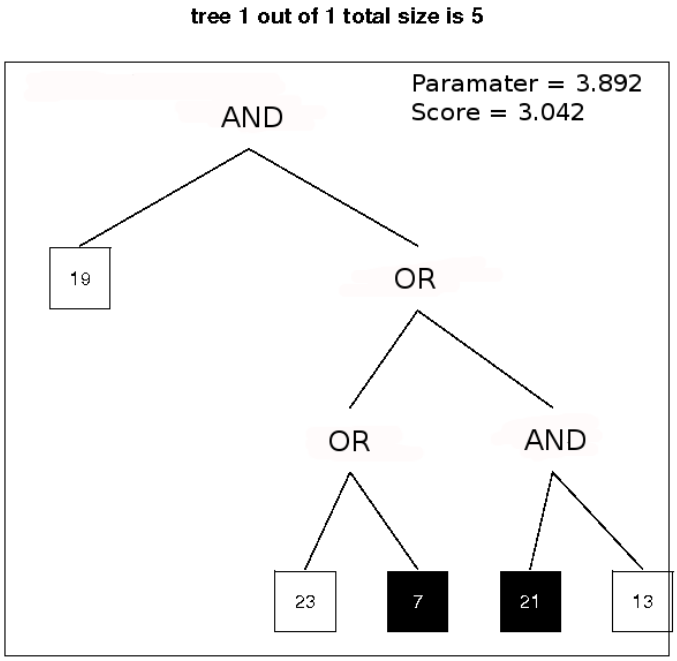


Figure 7.18: Comparison of outputs from raw logic regression step

MCMC Run		
Single	Double	Triple
19	17,13,23,21,14(MV),7	23,13,21,7,14(MV)
17	19,23,7,14(MV),13,21	23,13,21,7,14(MV)
23	7,21,17,19,1,14(MV)	21,17,19,13,1
13	19,17,23,14(MV),7,21	23,13,17,14(MV),3
7	23,17,13,21,19,10	21,17,19,13,7
21	23,13,21,17,14(MV),1	7,13,21,14(MV),17

Table 7.10: MCMC Comparisons from the most common variables derived from the five leaf models

tree level. In subfigure 7.17 it can be seen how variable 19 is AND'd with either variable 17 OR variable 23. This is extended in figure 7.18 where variable 19 is also AND'd with a host of variables and operators that include the significant variables 13 and 23.

It must also be noted that even though the reactive power variables were included in the logic model, only one of the reactive power variables was reported in the MCMC table 7.10. This G15 reactive power variable is shown in figure 7.19. It can be seen that G15 MVAr output is well enough correlated with the decay time constant profile toward the end of the sample window. This reduction in reactive power at generator

15 may indicate a drop in load in the network that affects the system state and reduces damping.

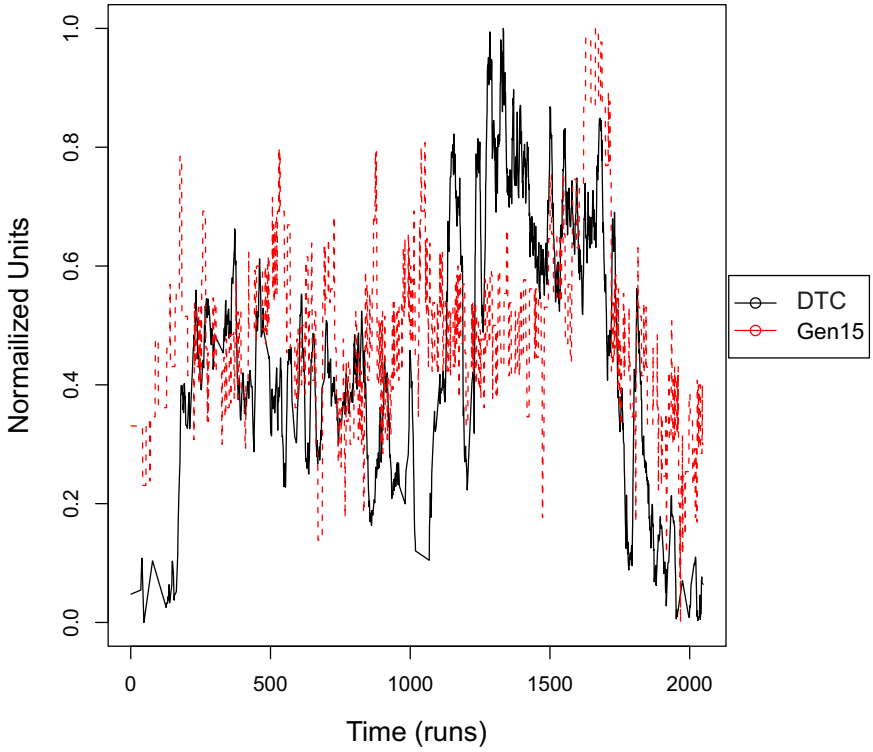


Figure 7.19: Comparison of 0.48Hz decay time constant and G15 MVAr plot

7.5.2 Validation of Raw Results

In this section, the reliability of the results from the previous section will be tested by comparing the generator outputs with the decay time constant of the 0.48Hz mode. Referring back to figures 7.17 and 7.18, the score and parameter values have been included for both the three and five leaf models. Both values are comparable to the values reported in the Icelandic examples in chapter 6 and in the 0.62Hz example earlier in sections 7.2.5 and 7.2.7, which indicates a well fitted model.

Like the WT logic models, there is little but some variation in the logic tree outputs in the three and five leaf cases i.e. the same trees are produced 90% of the time. However, the MCMC step still produces variant models which allows us to determine the groups of variables observed across the MCMC models.

In this section the three most prominent variables will be tested to determine if they are significant or not. This group consists of variables 17,19 and 23 (G20,G22,G34). In addition, some of the peripheral variables will also be tested as they have been reported as significant either directly or by association with significant variables.

Figure 7.20 shows the superposition of the 0.48Hz mode decay time constants against the generator outputs for G20,G22 and G34 as these were reported as the most significant variables from the logic model as well as the MCMC step. It can be seen how

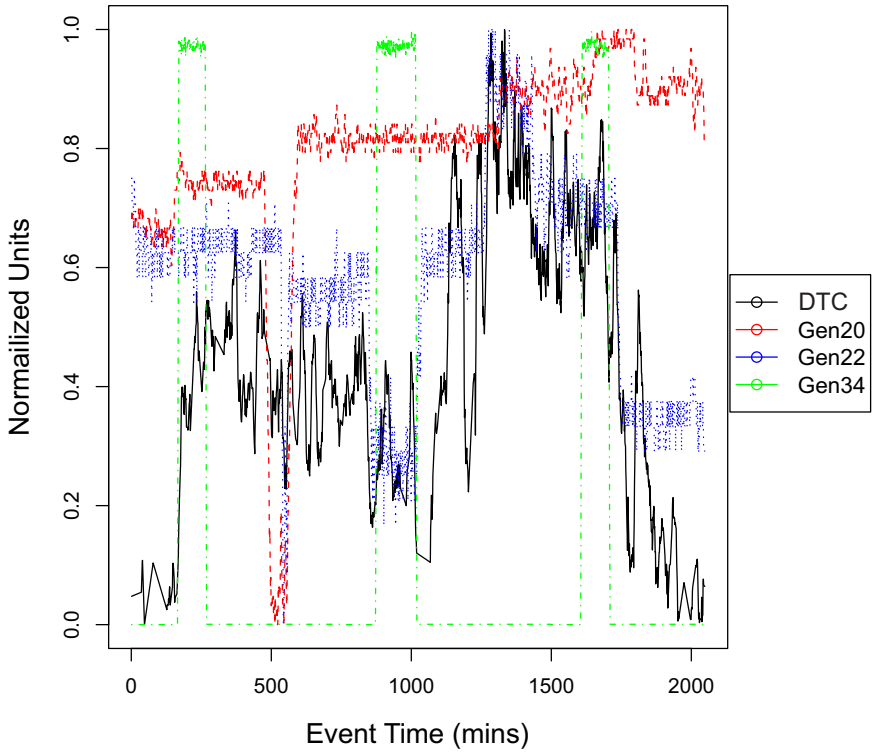


Figure 7.20: 0.48Hz Superimposed against G20,G22 and G34

the G20 and G22 outputs remain above the decay time constant plot for most of the window. G34 is a rather ambiguous case where it occasionally steps up but has no real relationship with the damping profile. This variable would be excluded in the WGLM model and as such wouldn't be considered or reported in a logic regression interaction. G22 was explained in section 7.3.3 using wavelets and was shown to be well correlated with the damping profile which suggests a strong relationship between them. Generators 20 and 34 have also been included as they both share "high" states during the poorly damped event.

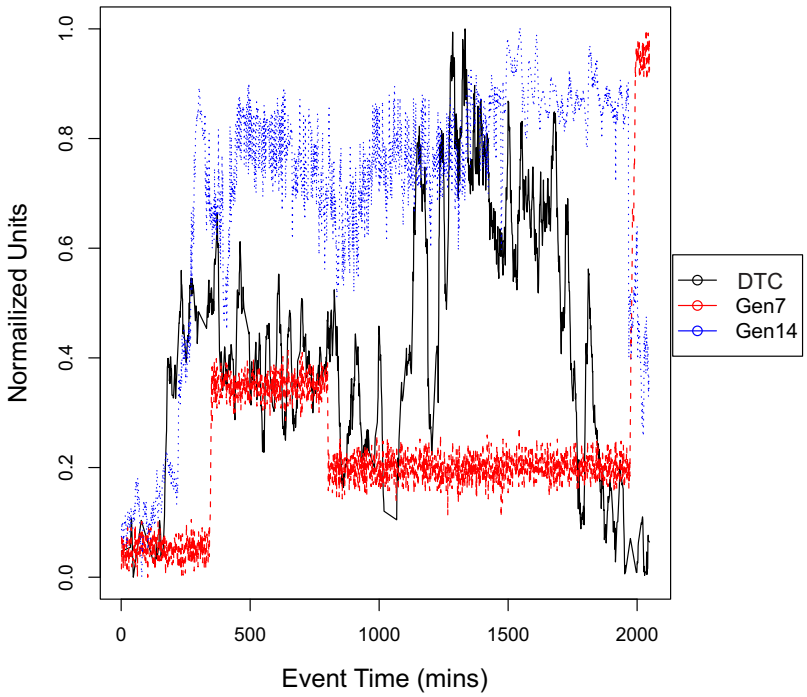
From the results presented in figure 7.20 it is reasonable to postulate that generators 20 and 22 act together to encourage poor damping. This conclusion can be made due to the fact that G20 is continually high for the decay time constant window and that G22 seems to track the decay time constant plot as well. In addition they are physically close to each other in the power system.

In figure 7.21, the subplots show the superposition of the 0.48Hz mode decay time constants against the generator outputs for G7,G14,G15 and G26. Subfigure 7.21(a) displays the relationships between the 0.48Hz mode damping and the G7 and G14 outputs. The G14 output is similar to the G1 output where both outputs are quite high with the G14 mean measured at 210MW. Previously, it was reported that the G7 output has a mean of 220MW so it is possible that the G7 output contributes to damping although it is not as prominent as the variables reported in figure 7.20.

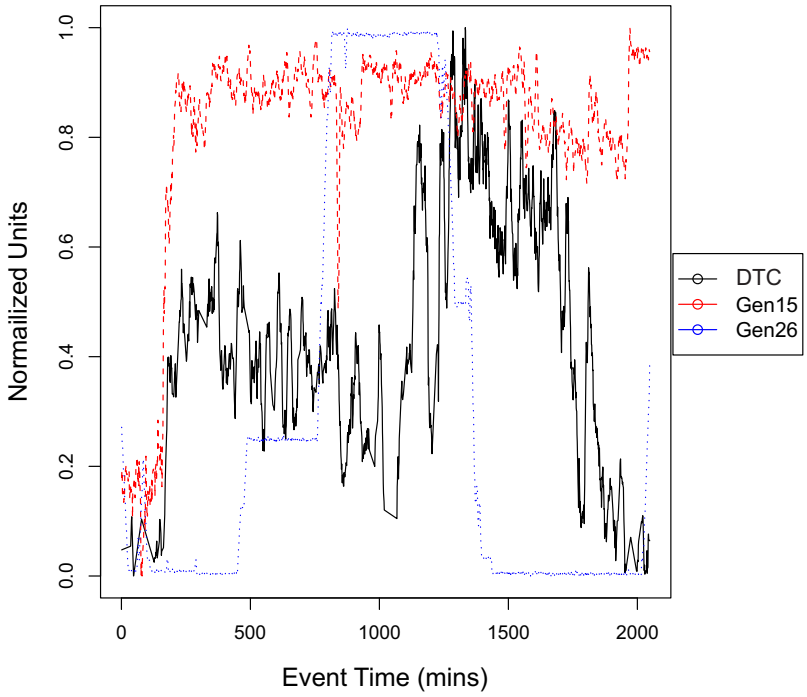
Subfigure 7.21(b) displays the relationship between the 0.48Hz mode damping and the G15/G26 outputs. Again, the G15 output is similar to the G1 and G14 outputs reported previously in that it goes high at the same time the poorly damped event begins. However, the average of the G15 output is approximately 70MW so it is debatable whether this machine has an effect on the mode damping, although it may act as a trigger for poorly damped events. G26, like G34 in figure 7.20, is ambiguous and it is difficult to see how it can have any effect on the mode damping. Again, this variable would be assigned a low WGLM coefficient and would be removed from any subsequent report. However, the G14 and G15 outputs have been reported before in section 7.3.5 and may be used to control mode damping with G22 being the most significant generator.

7.5.3 Results from other Australian Regions

In figure 7.22 the G25 (Queensland) output has been included as toward the end of the window the dip in the output coincides with the end of the poorly damped event. This generator located in Queensland has an average output of 420MW and so could well



(a) Gen7 and Gen15



(b) Gen15 and Gen26

Figure 7.21: Plots showing significant generator outputs against mode decay time constant

be involved with other machines in exerting an effect on mode damping. This variable was reported in interaction with generators 6 and 12 (Queensland) which are not well related to the mode damping. This is why G25 has been plotted alone.

Figure 7.23 displays the superposition of the T6 (InterArea Ties), the G12 (NSW)

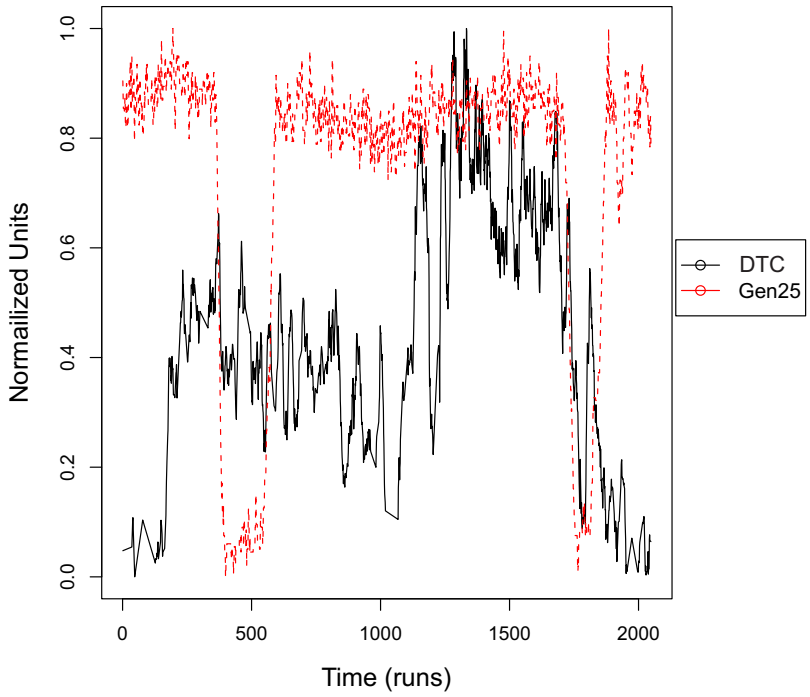


Figure 7.22: decay time constant vs. G25(Queensland)

MVAr outputs and the decay time constant profile which were reported in logic models derived from the Queensland and InterArea data.

In this case, G6 and G12 were reported as the high reactive power regions as well as their subsequent reductions compare well to the decay time constant profile. The T6 intertie flow relates to the link between NSW and Victoria, therefore the reduction in G12 (NSW) reactive power may reduce the intertie flow of reactive power. This change in reactive generation and flow into Victoria may be caused by a change in generation patterns in Victoria that has the effect of reducing the 0.48Hz mode decay time constants.

Although these reported variables have not been fully expressed in terms of their interactions they may still be pertinent in combination with other variables. As the methodology is improved upon, it should be possible to express these peripheral vari-

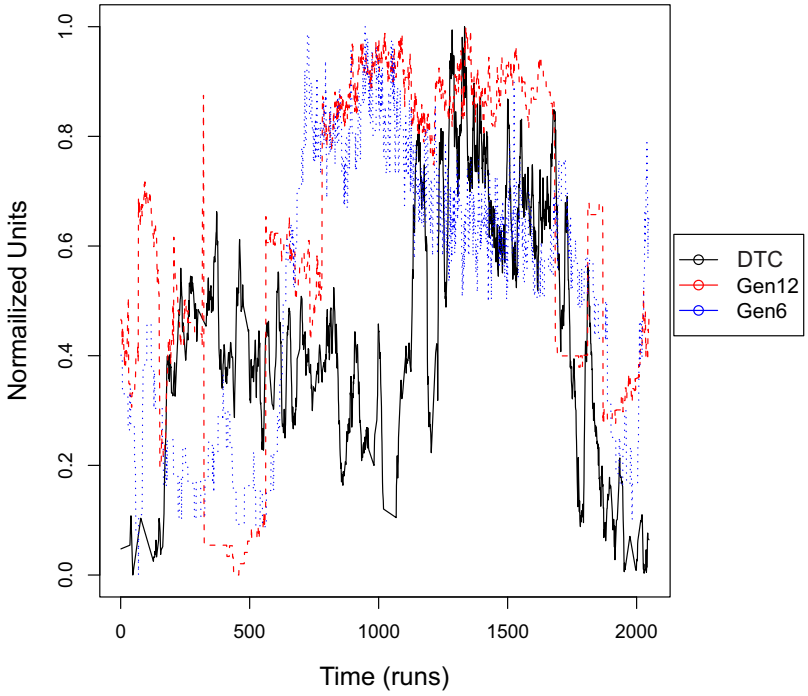


Figure 7.23: decay time constant vs. T6 and G12 from Queensland

ables more accurately in terms of their individual effects as well as their interactive effects with other variables.

7.6 Summary

Chapter 7 covered case studies including the 0.62Hz mode from the Icelandic model and a 0.48Hz mode from the real Australian system. The WGLM developed for the 0.62Hz Iceland case produced a model score of 72% with full variable selection and reactive power. By increasing the data length to 256, the model accuracy was increased to 77.5% which was accurate enough to extract significant parameters from the model that could be used as a preliminary study for the source location methodology.

The next step in chapter 7 was to develop full logic regression models from the wavelet transformed and raw data derived from the 0.62Hz Icelandic case. Just like the Icelandic 0.21Hz mode, it was discovered that both the raw and wavelet models overlapped to a great extent, although mutually exclusive variables did exist between the models. In the wavelet case, GB138 formed an interaction with GB126 and GB127 to compre-

hensively control the damping of the 0.62Hz mode.
In the raw case, GB139 formed an interaction with GB27 and GB29 to provide control over the mode damping. It was shown that either of these interactions could be used as a suitable control and it was suggested that both interactions could be amalgamated to increase the degree of damping control.

The Australian 0.48Hz mode was used to develop both wavelet transformed and raw logic regression models in an attempt to locate the source(s) of some resonant oscillations measured on the Australian system. The wavelet transformed (WT) logic model scores were 2.298 and 2.255 respectively for both three and five leaf models. These scores were comparable with the Icelandic 0.21Hz model which was well validated in chapter 6. In addition, this showed that the five leaf model was the optimum model and that the decay time constant dynamics could be explained using generation patterns or interactions containing five generators or less. From the WT model, generators 3, 15 and 22 were reported as significant and these were validated as far as possible by superposition of the MW outputs on the decay time constant profile.

The raw model produced similar results to the WT logic model. In this case, generators 20, 22 and 34 were reported by the logic model as well as variables T6 and G12 (both MVAr) from an InterTie and New South Wales (NSW) respectively. From the active power variables, generator 22 was validated via the WT logic model and G20 could be reasonably assumed to have an effect on mode damping. The reactive generation from InterTie T6 and G12 from NSW were shown to be linked and it was suggested they may have an influence on generation patterns in neighbouring Victoria, and that this could have an influence on the 0.48Hz mode damping.

Chapter 8

Conclusions and Further Work

8.1 Discussion

The results presented in chapter 5,6 and 7 show that the work undertaken confirms the usefulness of this approach in dealing with the oscillation source location problem. Variable selection was employed as a means of reducing the size and redundancy of the data set whilst not compromising model accuracy. This allowed a more compact data set to be entered into the wavelet and logic regression steps which increased the accuracy of the highlighted predictors.

By using the wavelet transform with the general linear model, one-way interactions between the response and predictor variables could be highlighted with less error in comparison to the regression models with raw variables detailed in the literature [45]. Logic regression was used on both raw and wavelet transformed data to develop models which could quantify the relationship between mode decay time constant and combinations of loadflow predictor variables. This was achieved by assigning regression coefficients to Boolean statements that contained these combinations. By highlighting combinations of generators that affect mode decay time, a more comprehensive and robust control of mode damping can be realized.

With the use of variable selection methods, wavelet transforms and logic regression, system operators will be able to identify the sources of electromechanical oscillations in

power systems in real time independently of an analytical dynamic model. Although, some historical data is needed initially to train the model, once a suitable number of data points have been collected (approximately 300), the algorithm could supply a system operator with a ranked list of loadflow variables that contribute to poor mode damping as well as information on how these loadflow variables interact in the system to produce poorly damped modes.
Current approaches have dealt with the source location problem using one-way correlation methods such as linear regression without dealing with the higher order and often more important effects related to generation patterns in the system that cause electromechanical modes to become poorly damped.

However, this measurement based approach will accompany model based approaches partly because of the success of the latter in terms of contingency analysis. With the increase of distributed generation and the uncertainty in load modelling the measurement based approach is becoming more important and combined with model based studies allows more informed choices to be made in determining suitable corrective control actions.

8.2 Conclusions

The work in this thesis showed that model and measurement-based stability analysis can reveal a great deal about the dynamic state of the network. New statistical techniques were applied to both dynamic models and real systems which yielded valuable information about the causes of the long decay time constants observed in these systems.
The measurement-based approach complements model-based dynamics analysis, rather than replacing it. Based on measurements, the behaviour of the system can be characterized using historical data with the aim of producing a real-time algorithm for source location. The main aim of this work was to identify modes close to stability limits and to determine contributing factors that produce the long decay terms observed. The contributing factors identified may take the form of single generator outputs or

interactions that define generation patterns in the system that are likely causes of poor damping of certain system modes.

In chapter 4, Constant admittance loads were discussed as a means to reduce the network size. The use of multiple loadflow scenarios along with random and intentional injection models were presented as a means of creating poorly damped events in the models. Random injections were limited to +/-20% nominal injection to simulate constant system activity and the intentional injections were used to ramp the generator outputs close to their limits. The resulting decay time constants, as well as the active and reactive power injections were used to develop statistical models to test the methodology.

In chapter 5, the 16 Machine and Icelandic System models were used and results were presented that supported the use of wavelet transformed general linear models (WGLM's) to locate the sources of oscillations. Based on some interesting results from the model, the results confirm the practical usefulness of this approach.
By wavelet transforming both the response and predictor variables, GLM's can be developed that can accurately predict the damping close to the system stability limits. This transformation caused the relationship between the damping and active/reactive power to become more linear which made them suitable for use in a GLM.

Using the Icelandic 0.21Hz mode and the corresponding system states, a WGLM was developed. Initially, 49% of the response variance was explained using the full frequency spectrum and this was increased to 73% with the application of optimal frequency scales. By incorporating both the reactive power and variable selection, the best scoring model was achieved with a 78.2% y-space variance accounted for. This meant that the regression coefficients derived from the model could be confidently used to control particular system modes.

This result was compared with the result from [45], where a score of 89% was reported from a standard multiple regression model. However, the same approach was applied

to the well validated Icelandic and 16 machine models which produced unsatisfactory results. It was also mentioned that the result from [45] was derived from a broader range of system variables including line flows and voltage bus magnitudes and angles. In addition, the model was developed using historical trends with predictors known to affect the 0.5Hz mode damping.

Chapter 6 introduced logic regression as a method of establishing important interactions in the power system that contribute to poor damping. The focus of the chapter was on the 0.21Hz Icelandic mode detailed in chapter 5. Both wavelet transformed and raw data were used to develop logic models in order to highlight significant interactions in the system.

Logic regression models were initially developed for wavelet transformed data with variable selection and optimal wavelet scales applied. GB27, GB99 and GB52 were shown to produce a significant interaction that could be used to explain poorly damped events. The wavelet transformed logic models agreed quite well with the WGLM and participation factors from chapter 5, however, some of the highlighted variables from the logic regression had diminutive participation factors. This was the case with GB52.
The raw data was then used to develop interactive logic models. The result was very similar to the wavelet transformed case in that GB27 and GB99 were reported as the most significant variables. The difference between the raw and wavelet models was highlighted by exclusive variables that were only reported in either of the models but not both. In the raw case, a five leaf model was shown to be optimal and a best scoring model of 2.566 was presented with a logic regression coefficient of 2.819. Interestingly, GB104 was highlighted as an important variable in the logic regression study but again, it had a negligible participation factor. By reducing the magnitude of combinations of generator outputs including GB104, it was shown that GB104 did have an effect on mode damping and that the differences between the statistical and analytical models were justified.

Reactive power was added to both the wavelet transformed and raw logic models

which had the effect of decreasing the model scores i.e. increasing the model accuracy. GB102/GB104 MVAr variables were highlighted in the logic regression analysis which was an interesting result as GB104 MW was also reported as significant in the active and reactive power models. The interaction of both active and reactive power in certain sections of the grid allowed a more detailed explanation of the mode dynamics to be developed.

Chapter 7 covered case studies including the 0.62Hz mode from the Icelandic model and a 0.48Hz mode from the real Australian system. The WGLM developed for the 0.62Hz Iceland case had a model score of 72% with full variable selection and reactive power included. By increasing the data length to 256, the model accuracy was increased to 77.5% which showed that increasing data lengths could produce better fitted models.

Full logic regression models for the wavelet transformed and raw data were derived from the 0.62Hz Icelandic case. Again, both the raw and wavelet models overlapped to a great extent, although mutually exclusive variables did exist between the models. In the wavelet case, GB138 formed an interaction with GB126 and GB127 which could be used to control damping. In the raw case, GB139 formed an interaction with GB27 and GB29 to provide control over the mode damping.

The Australian 0.48Hz mode was used to develop both wavelet transformed and raw logic regression models in an attempt to locate the source of the 0.48Hz oscillations. The wavelet transformed logic model scores were 2.298 and 2.295 respectively for both three and five leaf models. From the WT model, generators 3, 15 and 22 were reported as significant and these were validated as far as possible by superposition of the MW outputs on the decay time constant profile.

The raw model produced similar results to the WT logic model. In this case, generators 20, 22 and 34 were reported by the logic model as well as variables T6 and G12 (both MVAr) measured on an InterTie and in New South Wales (NSW) respectively. The reactive power from InterTie T6 and G12 from NSW were shown to be linked, and

it was suggested they may have an influence on generation patterns in neighbouring Victoria, and that this could have an effect on the 0.48Hz mode damping.
The case studies presented illustrate the use of the WGLM and logic regression approach both in terms of the model and the real system. The applications showed that a good prediction could be obtained, with most of the influential factors being captured in the statistical models.

In real-world applications, there will never be a perfect match between the statistical model behaviour and the observed behaviour. Discrepancies occur because:

- Variables influencing behaviour are not included in the data set
- Faults and failures in plant can change the dynamic behaviour
- Dynamics analysis inherently involves some scatter, due to non-white noise perturbation of the system

Some of the data available for the case studies was aggregated data, and this was used because detailed information was sensitive and unavailable for study purposes. PMU measurements contain much more detailed data, and it is expected that better statistical modelling could be achieved with a higher penetration of PMU's.

8.3 Further Work

The work reported shows that both model and measurement-based approaches to the identification of oscillatory instability is very promising. Both approaches use WGLM's as an initial step to highlight some of the main predictors in the statistical models as well as determining the relative importance of these variables in terms of their regression coefficients. Logic regression allows the combinatorial effect of the generators to be determined whilst the WGLM can simultaneously determine the most significant machines in a given combination.
It would be advantageous to develop the methods to produce automated tools to characterize the dynamic behaviour of the system and to investigate problems. This thesis

suggests that it would be feasible to develop such tools. Further work proposed is related to:

- Thresholding Techniques
- Fuzzy Logic for use in Logic Regression

8.3.1 Thresholding Techniques

It was repeadetdly stressed that the logic regression results were dependent on the thresholding used to dichotomize the response and predictor variables. At present, the thresholding is set as a function of variable means and variance and requires a certain amount of engineering judgment. However, an adaptive method including some form of feedback loop with model fits and scores would be preferable. By retroactively comparing the model fit and accuracy to the thresholds, it should be possible to find optimum thresholds for a given model and thus, the optimum results. This technique of adaptive filtering is already used in communications [81] where filter coefficients are constantly updated to adjust the frequency content of a signal. In the source location case, the filtering would be used to adaptively filter out low values thus retaining high ones.

8.3.2 Fuzzy Logic for use in Logic Regression

In [82], the logic regression methodology is taken one step further. Instead of a dichotomy of values (0 or 1) being used to represent the system, a number of intermediate steps are used to represent it. This method is known as fuzzy logic which has been reported in [1] where it was used to design a supervisory power system stabilizer (SPSS). In terms of the source location methodology, this could be applied to the response variable i.e. damping, to provide levels of alert. Instead of the logic regression reporting back combinations of variables for a mediocre decay time constant it would be possible to rank the severity of the alarm as well as the contribution of each predictor to the response variable. Only those variables assigned to the higher end of the fuzzy scale could be true predictors and the rest would be eliminated which would lead to a much more focused and accurate model. This fuzzyfication would be especially

useful in automating the process as it would drastically cut the number and complexity of the generator combinations reported.

8.4 Summary

The application of the source location methodology presented here for power system oscillations significantly extends the capability to understand dynamic behaviour. This leads to more effective defense against oscillatory instability, both in the operational and the planning context.

It was shown that a wide range of dynamic phenomena could be identified using the statistical method in both model and measurement-based systems. Also, statistical techniques could be used effectively to characterize the behaviour of the various observed dynamics in relation to the system state. Using case studies, a number of oscillation problems were illustrated, and it was shown that the source location techniques were practically useful [2].

The work presented in this thesis is by no means complete as further work has to be done to determine the robustness of the method. As a result, the algorithm has to be tested on a greater variety of modelled and real data to access if it performs well in situations where the generation mix changes significantly, there are changes in load profile and where there is an increased penetration of FACTS devices.

With further testing completed, it is envisioned that the source location algorithm will be written into Psymetrix'x PhasorPoint platform. By using PhasorPoint mode dynamics output (mode decay time constant, phase, amplitude and frequency) in conjunction with an energy management system (EMS), the source location algorithm can be used to locate variable sources of electromechanical oscillations across the entire power system. This will allow system operators to take well informed action in real time, to alleviate damping problems in the power system before any components are damaged.

Bibliography

[1] H. behbehbani, *"Supervisory PSS"*. PhD Thesis, University of Edinburgh, 2008

[2] Douglas H. Wilson, *"Identification of Causes of Poorly Damped Oscillations Observed in Power Systems"*. MSc Thesis, UMIST, 2006

[3] Lee, D.C., Beaulieu, R.E., and Service, J.R.R., *"The Design of Experiments"* IEEE Trans. Power Appar. and Sys., 100:99, 4151-4157, 1981.

[4] P.Kundur, *"Power System Stability and Control"*. McGraw-Hill, 1994

[5] G.Rogers, *"Power System Oscillations"*. Kluwer, 2000

[6] Hauer J. F., Demeure C.J., Scharf L.I., *"results in Prony analysis of power system response signals"* IEEE Trans Power Systems, Vol.5, No.1, pp80-89, 1990, 4151-4157, 1981.

[7] E.V. Larsen and D.A. Swann, *"Applying power system stabilizers"*. IEEE Trans. PAS-100 (1981), pp. 30173046

[8] CIGRE Task Force 38.01.07, *"Analysis and Control of Power System Oscillations"*. Brochure 111, December 1996

[9] Erinmez I.A., Humphreys P., Geeves S.S, *"Application of power system stabilisers on the Anglo-Scottish interconnection. Development of analytical techniques for system damping evaluation"*. IEE Proc. Generation, Transmission and Distribution, Vol 135, Iss 3, May 1988

[10] Kosterev D.N., Taylor C.W., Mittelstadt W.A., *"Model Validation for the August 10, 1996 WSCC Outage"*. IEEE Trans. Vol 14, No 3, August 1999

[11] NERC Disturbance Analysis Working Group, "DAWG Database" *"http://www.nerc.com/ dawg/database.html "*

[12] Venkatasubramanian V., Li Y, *"Analysis of 1996 Western American Electric Blackouts"*. Proc. IREP conference, Italy, Aug '04

[13] Hiskens I. A., Akke M, *"Analysis of the Nordel Power Grid Disturbance of January 1, 1997 Using Trajectory Sensitivities"*. IEEE Trans Power Systems, Vol. 14, No. 3, August 1999

[14] Power Failure in Eastern Denmark and Southern Sweden on 23 September 2003, Preliminary report, Elkraft System, 2 October 2003

[15] L. Wehenkel, *"Automatic Learning Techniques in Power Systems"*. Kluwer Academic Publishers, 1998

[16] M.B. Zayan, M.A. El-Sharkawi, N.R. Prasad. *"Comparative Study of Feature Extraction Techniques for Neural Network Classifier"*. Proc. Conf. Intelligent Systems Applications to Power Systems, ISAP '96, pp. 400-404, 1996.

[17] C.A Jensen, M.A. El-Sharkawi, R.J. Marks II, *"Power System Security Assessment of Neural Networks using Fisher Discrimination"*. IEEE Trans. on Power Systems, vol.16, no.4, November2001

[18] I.N. Kassabalidis, M.A. El-Sharkawi, *"Dynamic Security Border Identification using Enhanced Particle Swarm Optimization"*. IEEE Trans. on Power Systems, vol.17, no.3, August 2002.

[19] L.S. Moulin, M.A. El-Sharkawi, R.J. Marks, A.P. Alves da Silva. *"Automatic Feature Extraction for Neural Networks Based Power Systems Dynamic Security Evaluation"*. ISAP 2001, no.21, Budapest, June 2001.

[20] R.O. Duda. *"Pattern Classification"*. New York, Wiley, 2001.

[21] S. Teeuwsen, *"Oscillatory Stability Assessment of Power Systems using Computational Intelligence"*. Shaker-Verilag, ISBN 3-8322-4037-3

[22] I.T. Joliffe, *"Principal Component Analysis"*. Shaker-Verilag,

[23] Visual Numerics, IMSL Fortran 90 MP Library Help *"Stat/Library"*. Vol.2, Chap.11, Cluster Analysis.

[24] A. Brieman, *"Classification and Regression Trees"*. Wadsworth. Int, California. 1984

[25] R. Jang, *"Neuro-Fuzzy and Soft Computing"*. Prentice Hall, NJ. 1997

[26] J.R. Quinlan, *"Induction of Decision Trees"*. Machine Learning, Vol.1, pp.81-106, 1986

[27] J.R. Quinlan, *"C4.5 Programs for Machine Learning"*. Morgan Kauffman, San Mateo, California, 1993

[28] C. Huang, J.R.G. Townsend, *"A Stepwise Regression Tree for NonLinear Approximation: Applications to Esimating Subpixel Land Cover"*. International Journal of Remote Sensing, Vol.24, No.1, 75-90, 2003.

[29] H. Kim, G.J. Koehler, *"Theory and Practice of Decision Tree Induction"*. Omega, Vol.23, No.6, pp.637-652,1995

[30] K.F. Man, K.S. Tang, *"Genetic Algorithms"*. Springer, London 1999

[31] L.J. Cai, *"Robust Coordinated Control of FACTS Devices in Large Power Systems"*. PhD Thesis, University of Duisberg-Essen. February 2004.

[32] M.A. El-Sharkawi, S.J. Huang, *"Ancillary Technique for Neural Network Applications"*. IEEE Int. Conf. Neural Networks, Orlandi, Fl. pp. 3724-3729, June 1994

[33] M.A. El-Sharkawi, *"Neural Networks Power"*. IEEE Potentials, Vol.15, no.5, pp. 12-15 December 1996.

[34] Y. Mansour, E. Vaahedi, M.A. El-Sharkawi, *"Dynamic Contingency Screening and Ranking using Neural Networks"*. IEEE Trans. on Neural Networks, Vol.8, no.4, pp. 942-950. July 1997.

[35] Fuzzy Logic Toolbox *"For use with Matlab"*. The Mathworks Inc. 1999.

[36] R.D. Reed, R.J. Marks II, *"Neural Smithing"*. MIT Press, Cambridge, Mass, 1999.

[37] S.S. Haykin, *"Neural Smithing"*. MIT Press, Cambridge, Mass, 1999.

[38] C.M. Bishop, *"Neural Networks for Pattern Recognition"*. Clarendon Press, Oxford, 1996.

[39] Neural Network Toolbox *"For use with Matlab"*. The Mathworks Inc. 2001.

[40] R. Fuller, *"Introduction to Neuro-Fuzzy Systems"* Pysica-Verlag Heidelberg,2000

[41] E. Czogala, *"Fuzzy and Neuro-Fuzzy Intelligent Systems"* Physica-Verlag, Heidelberg,2000

[42] H.H. Boethe, *"Neuro-Fuzzy"* Springer, Berlin, 1998

[43] T. Takagi, M. Sugeno, *"Fuzzy Identifications of Systems and its Applications to Modelling and Control"*. IEEE Trans. Syst, Man., Cybern. Vol.15, no.4, pp. 116-132. July 1985.

[44] E. Lughofer, *"Online Adaptation of Takagi-Sugeno Fuzzy Interference Systems"* Technical Report FLLL-TR-0217, Fuzzy Logic Laboratory, Linz-Hagenberg

[45] Archer. B.A, Annakage. U.D, Wijetunge. P, *"Accurate Prediction of Damping in Large Interconnected Power Systems with the aid of Regression Analysis"*. IEEE Trans. Power Sys., 23, 3, 1170-1178 2008

[46] Fisher. R.A, *"The Design of Experiments"* Hafner, 1966 ISBN-10: 0028446909

[47] Lubosny. Z, *"Wind Turbine Operation In Electric Power Systems: Advanced Modeling"*. Spingler, 2003

[48] Nomikos. B.M, Kotlida. M.A, Vournas. C, *"Interarea Oscillations and Tie-Line Transients in the Hellenic Interconnected System"*. PowerTech 2007, 2007

[49] Machowski. J, Bialek. J, Bumby. J, *"Power System Dynamics and Control"*. Wiley, 2008

[50] McNabb. P, Wilson. D, Bialek. J, *"Dynamic Model Validation of the Icelandic Power System using WAMS based Measurements of Oscillatory Stability"* PSCC, Glasgow, 2008

[51] Wilson. D, Hay. K, McNabb. P, *"Identifying Sources of Damping Issues in the Icelandic Power System"*. IEEE.InProceedings, PSCC, Glasgow, 2008.

[52] Anaparthi, K.K *"Measurement Based Identification and Control of Electromechanical Oscillations in Power Systems"* IEEE Power India Conference, 10 - 12 April 2006, New Delhi 2006

[53] B. Jayasekara, *"Derivation of an Accurate Polynomial Representation of the Transient Stability Boundary"*. IEEE Trans. Power Sys., 21, 4, 1856-1863 2006

[54] Meyer. Y, *"Wavelet Algorithms and Applications"*. Academic Press, 1994 ISBN-10: 0898713099

[55] Cohen. A, *"Non-Stationary Multiscale Analysis"*. Wavelets: Theory and Applications, pp. 3-12, ISBN 0-12-174575-9

[56] Chui. C, *"Wavelet Theory and Applications"*. Technology & Engineering, ISBN-10: 0471197483

[57] Daubechies. I, Lagarias. J, *"Two-Scale Difference Equations"*. SIAM Journal on Mathematical Analysis, 23, 4, 1031-1079

[58] Crandall. R, *"Projects in Scientific Computing"*. Telos, ISBN-10: 0387978089

[59] Amara. R, *"An Introduction to Wavelets"*. Computational Science & Engineering, IEEE, 2, 2, 50-61, 211-212.

[60] Burrus. C, Gopinath. R, Guo. H, *"Introduction to Wavelets and Wavelet Transforms: A Primer"*. 1998, Upper Saddle River, NJ (USA): Prentice Hall,

[61] Calderbank. A.R, Daubechies. I, Sweldens. W, Yeo. B.L, *"Wavelet Transforms that map integers to integers"*. Proceedings of the IEEE Conference on Image Processing. Preprint, IEEE Press, 3, 906-966.

[62] Abraham. B, Ledolter. J, *"Statistical Methods for Forecasting"*. New York, Wiley, 2005 ISBN-10: 0471867640

[63] Miller. I, Freund. J.E, *"Probability and Statistics for Engineers"*. New Jersey, Prentice-Hall, 1985 ISBN-10: 0138402086

[64] Benoudjit. N, Franois. D, Meurens. M, Verleysen. M, *"Spectrophotometric Variable Selection by Mutual Information"* Chemometrics and Intelligent Laboratory Systems, 74, 2, 243-251 2004

[65] Abdi. H, *"Partial Least Squares Regression"* Encyclopedia for research methods for the social sciences, Thousand Oaks (CA) 792-795

[66] Shannon. C.E, *"A Mathematical Theory of Communication"* American Telephone and Telegraph Co, New York 463-479 1948

[67] Battiti. R, *"Using Mutual Information for Selecting Features in Supervised Neural Net Learning"* IEEE Trans. Neural Networks, 5, 537-550 1994

[68] Beardah. C.C, Baxter. S.M.J, *"Computer Applications and Quantitive Methods in Archaelogy"* 1998 British Archaeological Reports, International Series, 757-767 1995

[69] Alsberg. B, Woodward. W.M, Winson. M.K, Rowland. J.J, Kell. D.B, *"Variable Selection in Wavelet Regression Models"* Analytica Chimica Acta, 29-44 1998

[70] Nelder. J, Wedderburn. R, *"Generalized Linear Models"* Journal of the Royal Statistical Society, 135, 3, 370-384 1972

[71] McCullagh. P, Nelder. J, *"Generalized Linear Models"* Chapman and Hall, 1989 ISBN 0-412-31760-5

[72] Cardoso. J, *"Informax and Maximum Likelihood for Source Seperation"* IEEE Letters on Signal Processing, 4, 3, 112-114. 1997

[73] Mardia. K.V, Kent. J.T, Bibby. J.M, *"Multivariate Analysis"* Academic Press, 1979 ISBN 0-12-471252-5

[74] Cox. D.R, Snell. E.J, *"Applied Statistics: Principles and Examples"* Chapman and Hall, 1981 ISBN 0-412-16570-8

[75] Hastie. T.J, Tibshirani. R.J, *"Generalized Additive Models: Principles and Examples"* Chapman and Hall, 1981 ISBN 9780412343902

[76] Carroll. R.J, Ruppert. D, *"Transformation and Weighting in Regression"* Chapman and Hall, 1988

[77] Saltelli, A., S. Tarantola, and K. Chan, *"Quantitative model-independent method for global sensitivity analysis of model output"* Technometrics, 41, 1, 39-56 2004

[78] R.J. Latino and K.C. Latino, *"Root Cause Analysis Improving Performance for Bottom Line Results"* CRC Press, ISBN-13: 9780849353406

[79] Bhat. A, Lucek. P, Ott. J, *"Analysis of a Complex Disease using Artificial Neural Networks"* Genetic Analysis Workshop 11, Arcachon, France, 1998

[80] Green. P.J *"Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination"* Biometrika, 82, 4, 711-732

[81] Mulgrew. B, Grant. P, Thompson. J, *"Digital Signal Processing: Concepts and Application"*. Academic Press, 2002

[82] Kooperberg. C, Ruczinki. I *"Identifying interacting SNPs using Monte Carlo Logic Regression"* Genetic Epidemiology, 2005

[83] Grossman. A, Morlet. J, *"Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape"*. 1984, SIAM J.Math.Anal, 15, 723-726

[84] Torrence. C, Compo, G.C, *"A Practical Guide to Wavelet Analysis"*. Bulletin of the American Meteorological Society, 79, 61-78, 2005

[85] Acharya. R, Suri. S, Spaan, J.A, *"Advances in Cardiac Signal Processing"*. Springer, 2007, ISBN: 3540366741

[86] Sheng. Y, *"Wavelet Transforms"*. Transforms and Applications Handbook, Boca Raton, Fl (USA): CRC Press, 1996, 747-827

[87] Daubechies. I, *"Ten Lectures on Wavelets"*. CBMS-NSF regional conference series in applied mathematics 61, Philadelphia: SIAM, 1992, ISBN-10: 0898712742

[88] Steeb. W.H, *"Non-Linear Workbook"*. World Scientific Publishing Company, 2005, ISBN-10: 9812562788

[89] Daubechies.I, *"Orthogonal Bases of Compactly Supported Wavelets"*. Comm Pure & Appl. Math, 41, 909-996 1988

[90] Press. W, *"Numerical Recipes in Fortran"*. Cambridge University Press, New York 1992

[91] Cabrelli. C, Molter, U, *"Generalized Self Similarity"*. Journal of Mathematical Analysis and Applications, 230, 251-260. 1999

[92] Kammler. D.W, *"A First Course in Fourier Analysis"*. Prentice-Hall, Upper Saddle River, NJ, 2008 ISBN 0-13-578782-3

[93] Mallat. S.G, *"A Theory for Multiresolution Signal Decomposition: The Wavelet Representation"*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 7, 674-693. 1989

[94] Chen. W.K, *"The Circuits and Filters Handbook"*. A CRC Handbook. Published in Cooperation with IEEE, 2002 ISBN-10: 0849383412

[95] EPRI, *"A Dynamic Information Manager for Networked Monitoring of Large Power Systems"*. Palo Alto, CA, 1999 ISBN-10: 0849383412

[96] Hauer. J.F, DeSteese. J.G, *"An Oscillation Detector and Characterization of Special Behaviour in Large Electric Power Systems"*. IEEE Trans on Power Systems, 22, 123-134, 2004

[97] Schulz. R.P, Laios. B.B, *"Triggering Trade-Offs for Recording Dynamics"*. IEEE Trans on Power Systems, 10, 44-49 1997

[98] Mei. K, Rovnyak. S.M, Ong. C.M, *"Dynamic Event Detection using Wavelet Analysis"*. Power Engineering Society General Meeting, 2006. IEEE, 10, 7-14 2006

[99] Bickel. P.J, *"Frontiers in Statistics"*. Imperial College Press, 2006

[100] Donohoe. D, Johnstone. J, *"Ideal Spatial Adapdation by Wavelet Shrinkage"*. Biometrika, 81, 1, 425-455 1992

[101] Abramovich. F, Benjamini. Y, *"Adaptive Thresholding of Wavelets"*. Journal of Computational Statistics & Data Analysis, 22, 351-361 1996

[102] Nason. G.P, *"Wavelet Shrinkage using Cross-Validation"*. J. R. Statist. Soc. B., 58, 463-479 1996

[103] Cai. T.T, *"Adaptive Wavelet Estimation: A Block Thresholding Approach"*. Annals of Statistics, 58, 463-479 1999

[104] Givant. S, Halmos. P *"Introduction to Boolean Algebras"* Springer Verlag, 2008 ISBN 978-1-59102-089-9

[105] Breiman. L, Friedman. J, Stone. C.J, Olshen. R.A *"Classification and Regression Trees"* Wadsworth Co, 1995 ISBN-10: 0412048418

[106] Chipman. H, George. E, McCulloch. R, *"Bayesian CART Model Search"* Journal of the American Statistical Association, 93, 3, 935-960 1998

[107] Kirkpatrick. S, Gelatt Jr. C.D, Vecchi. M.P, *"Optimization by Simulated Annealing"* Annals of Science, 220, 4598, 671-680 1983

[108] Otten. R, Van Ginneken. L *"The Annealing Algorithm"* Kluwer, 1998 ISBN-10: 0792390229

[109] Laarhoven. P.J, Aarts. E.H *"Simulated Annealing: Theory and Applications"* Springer Verlag, 1987 ISBN 90-277-2513-6

[110] Fritsch. A, Ickstadt. K *"Comparing Logic Regression Based Methods for Identifying SNP Interactions"* Lecture Notes in Computer Science, Springer 2007

[111] Ruczinski. I, Kooperberg. C, LeBlanc. M *"Logic Regression"* Journal of Computational and Graphical Statistics, 12, 3, 475-511 2003

[112] Ruczinski. I, Kooperberg. C, LeBlanc. M *"Polychotomous Regression"* Journal of the American Statistical Association, 90, 78-94 1995

[113] M. Jansen, R. Baraniuk, and S. Lavu *"Multiscale Approximation of Piecewise Smooth Two-Dimensional Functions using Normal Triangulated Meshes"* Appl. Comp. Harm. Anal., 19(1), pages 92-130, 2005.

[114] M. Holschneider *"Wavelet analysis on the circle"* J. of Math. Phys. , 31(1):39-44, January 1990, 19(1), pages 92-130, 2005.

[115] I. Daubechies *"Orthonormal bases of compactly supported wavelets"* Comm. on Pure and Appm Math, 41(7), pages 909-996, 1988.

Appendix A

Wavelet Transform

A.1 Continuous Wavelet Transform

The continuous wavelet transform was introduced in 1984 by Morlet and co-workers to analyze geophysical signals with some kind of modified windowed Fourier transform (WFT), which reads [113]:

$$\mathcal{F}[s](t,\omega) = \frac{1}{\sqrt{2\pi}} \int_R s(y)\overline{h(y-t)}e^{-i\omega y}dy \tag{A.1}$$

with $s \in L_2(\mathrm{R})$ being a time-continuous signal and $h \in L_2(\mathrm{R})$ a window function. Here $L_2(\mathrm{R})$ denotes the space of square integrable functions on R. The modification of the WFT was established by combining the window function and Fourier mode $e^{-i\omega y}$ into one window function ψ that can be scaled.

A continuous wavelet transform is used to divide a continuous-time function into wavelets. Unlike the Fourier transform, the continuous wavelet transform possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization. A wavelet transform is a convolution of a signal $x(t)$ with a set of functions which are generated by translations and dilations of a main function. The main function is known as the mother wavelet and the translated or dilated functions are called daughter wavelets. Mathematically, the continuous wavelet

transform (CWT) is given by [83]:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\psi^* \left(\frac{x-b}{a} \right) dx \tag{A.2}$$

where ψ^*(x) is the analyzing or mother wavelet function and $\psi \in L_2(\mathrm{R})$, $a(>0)$ is the scale parameter with $b \in \mathrm{R}$ the position or translation parameter.
The wavelets are generated from a single *mother* wavelet $\psi(t)$ by scaling and translation:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{s} \right) \tag{A.3}$$

The transform is characterized by the following properties:

- It is covariant under translations
- It is covariant under dilations

These properties make the wavelet transform very suitable for analyzing hierarchical structures and can be given mathematically as:

$$f(x) \to f(x-u) \qquad W(a,b) \to W(a,b-u) \tag{A.4}$$

$$f(x) \to f(sx) \qquad W(a,b) \to s^{-\frac{1}{2}} W(sa,sb) \tag{A.5}$$

The dilation property is especially useful as the wavelet transform is, in essence, a mathematical microscope with properties that do not depend on magnification.

The analyzing wavelet function ψ(x), also known as the 'mother wavelet' is continuous in both the time and the frequency domains. The main purpose of the mother wavelet is to provide a source function to generate the daughter wavelets, which are simply the translated and scaled versions of the mother wavelet.

In Fourier space, we have:

$$\hat{W}(a,v) = \sqrt{a}\hat{f}(v)\hat{\psi}^*(av) \tag{A.6}$$

When the scale a varies, the filter $\hat{\psi}^*(av)$ is only reduced or dilated whilst maintag the overall wavelet pattern.

Now consider a function W(a,b) which is the wavelet transform of a given function f(x). It has been shown by [114] that f(x) can be restored using the formula:

$$f(x) = \frac{1}{C_\chi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{a}} W(a,b) \chi(\frac{x-b}{a}) \frac{da.db}{a^2} \tag{A.7}$$

where:

$$C_\chi = \int_0^{+\infty} \frac{\hat{\psi}^*(v)\hat{\chi}(v)}{v} dv = \int_{-\infty}^{0} \frac{\hat{\psi}^*(v)\hat{\chi}v}{d} v \tag{A.8}$$

Generally χ(x) = ψ(x), but other choices can enhance certain features for some applications. The reconstruction is only available if C_ψ is defined (admissibility condition). In the case of χ(x) = ψ(x), this condition implies $\hat{\psi}(0) = 0$. i.e. the mean of the wavelet function is zero.

A.1.1 Mother Wavelet

In wavelet theory, the scaling and translation operators act simultaneously on the mother wavelet function. Performing affine operation on the mother wavelet creates a set of scaled translated copies of the original mother wavelet. This is called a wavelet set and the mother wavelet is called the kernel of the wavelet transform. An example of a mother wavelet that is used to represent a given signal at different scales and resolutions is shown in figure A.1. The CWT of a function $f(x)$ was given in equation (A.2). In general, it is preferable to choose a mother wavelet that is continuously differentiable with a compactly supported scaling function and high vanishing moments i.e. the Fourier transform of ψ(t) vanishes at zero frequency. A wavelet is defined by the following two functions: the wavelet function ψ(t) and the scaling function φ(t).

We seen how a function 'f' in 'x' was transformed into another function 'a' in 'b'. A wavelet coefficient $W_\psi f(a,b)$ at a particular scale and translation represents how well the signal 'f' and the scaled and translated mother wavelet match; or the coefficient represents the "degree of correlation" between the functions at a particular scale and translation. The conditions that need to be satisfied for our function ψ(t) to behave as

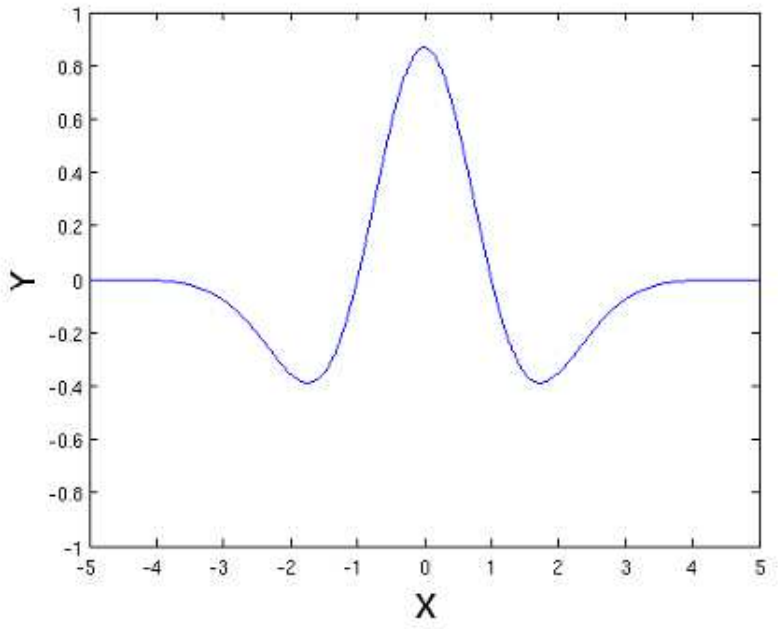


Figure A.1: Mexican Hat Wavelet often used as a Mother Wavelet in analyzes

a wavelet are summarized below for convenience [84]:

$$\int \psi(t)dt = 0 \tag{A.9}$$

$$\int |\psi(t)|^2 \, dt = 0 \tag{A.10}$$

$$C_\chi = \int_{-\infty}^{+\infty} |\psi(\omega)|^2 \, / \, |\omega| \, d\omega \; < \infty \qquad C_\chi \neq 0 \tag{A.11}$$

If we define $\phi_{a,b}$(t) as the translated and scaled version of ϕ(t), then:

$$\psi_{a,b} = (1/\sqrt{|a|})\psi((t-b)/a) \tag{A.12}$$

The multiplication with $1/\sqrt{|a|}$ is to satisfy the second condition given in equation (A.10). (Normalizing $\rightarrow \|\psi\| = 1$).

So $W_\psi f(a,b) = \langle f(t), \psi_{a,b}(t) \rangle$ with CWT as the inner product. The concept of CWT is shown in figure A.2 below [85].

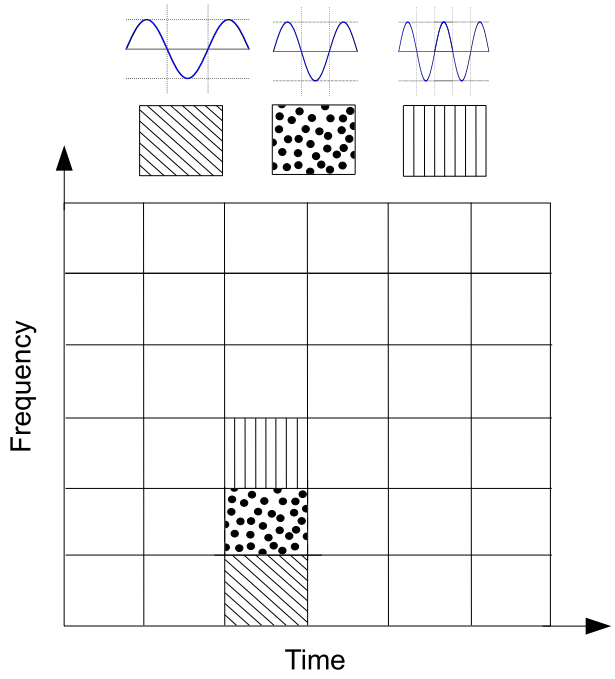


Figure A.2: Time Frequency Tiling

A.1.2 Scaling Function and Scale Factor

The wavelet function $\phi(t)$ and the scaling function $\varphi(t)$ define a wavelet. The scaling function is primarily responsible for improving the coverage of the wavelet spectrum. This could be difficult since time is inversely proportional to frequency. In other words, if we want to double the spectrum coverage of the wavelet in the time domain, we have to sacrifice half the bandwidth in the frequency domain. Instead of covering all the spectrum with an infinite number of levels, a finite combination of the scaling function is used to cover the spectrum. As a result, the number of wavelets required to cover the entire spectrum has been greatly reduced.

The scale factor a either dilates or compresses a signal. When the scale factor is relatively low, the signal is more contracted which in turn results in a more detailed composition. However, the drawback is that low scale factor does not last for the duration of the signal. On the other hand, when the scale factor is relatively high, the signal is stretched out which means the resulting composition will be presented in less detail. Nevertheless, it usually lasts the entire duration of the signal.

A.1.3 Wavelet Properties

As evident in equation (A.2) the wavelet transform of a one-dimensional function is two-dimensional; the wavelet transform of a two-dimensional function is four dimensional. The time-bandwidth product of the wavelet transform is the square of the input signal and for most applications this is not a desirable property. Therefore, some additional conditions have to be imposed on the wavelet functions in order to make the wavelet transform decrease quickly with decreasing scales. These are the *regularity conditions* and they state that the wavelet function should have some smoothness and concentration in both time and frequency domains. Regularity is quite a complex concept but can be explained using the concept of *vanishing moments*.

If we expand the wavelet transform in equation (A.2) into the Taylor series at t=0 until order n (let $\tau = 0$ for simplicity) we get [86]:

$$\gamma(a,0) = \frac{1}{\sqrt{a}} \left[\sum_{p=0}^{n} f^p(0) \int \frac{x^p}{p!} \psi\left(\frac{x}{a}\right) dt + O(n+1) \right] \tag{A.13}$$

Here $f^{(p)}$ stands for the p^{th} derivative of f and $O(n+1)$ denotes the rest of the Taylor expansion. If we now define the *moments* of the wavelet by M_p:

$$M_p = \int x^p \psi(x) dt \tag{A.14}$$

then we can rewrite equation (A.13) in to the finite development:

$$\begin{aligned} \gamma(a,0) \quad = \quad & \frac{1}{\sqrt{a}} \left[f(0) M_0 a + \frac{f^{(1)}(0)}{1!} M_1 a^2 + \frac{f^{(2)}(0)}{2!} M_2 a^3 + \dots \right. \\ & \left. \dots + \frac{f^{(n)}(0)}{n!} M_n a^{n+1} + O(a^{n+2}) \right] \end{aligned}$$

From the admissibility condition we already have the 0^{th} moment M_0, so that the first term on the right-hand side of equation (A.15) is zero. If we now make the other moments up to M_n equal to zero as well, then the wavelet transform coefficients $\gamma(a,b)$

will decay as fast as s^{n+2} for a smooth signal $f(x)$. This is known in the literature as the vanishing moments or the *approximation order*. If a wavelet has N vanishing moments, then the approximation order of the wavelet transform is also N. The moments do not have to be exactly zero, a small value is often good enough. In fact, experimental research suggests that the number of vanishing moments required depends heavily on the application [61].

In summary, the admissibility condition gave us the wave, regularity and vanishing moments gave us the fast decay, and combined they give us the wavelet. More about regularity and vanishing moments can be found in the literature [60] [87].

A.2 Discrete Wavelet Transform

In the previous section the components that combine to produce the wavelet transform were discussed. It is important at this stage to make the wavelet transform more practical for the source location application. However, the wavelet transform as described thusfar still has three properties that make it difficult to use directly in the form of equation (A.2). The first of these being the redundancy associated with the CWT. In equation (A.2) the wavelet transform is calculated by continuously shifting a scalable function over a signal and calculating the correlation between the two. It is clear that these scaled functions will not adhere to an orthogonal basis and as a result, the obtained wavelet coefficients will be highly redundant.

The CWT behaves just like an orthogonal transform in the sense that the inverse wavelet transform permits us to reconstruct the signal by an integration of all the projections of the signal onto the wavelet basis. This is called quasi-orthogonality [86].For most practical applications it would be useful to remove this redundancy.

Even without the redundancy of the CWT, the number of wavelets in the wavelet transform is still infinite and this would need to be finite to make the transform practical. This condition forms the second obstacle in achieving a practical wavelet transform.

The third problem is that for most functions the wavelet transforms have no analytical

solutions and can only be calculated numerically or by an optical analogue computer. Fast algorithms are needed to exploit the power of the wavelet transform and it is the existence of the these algorithms that have made the wavelet transform a useful engineering concept.

As mentioned before the CWT maps a one-dimensional signal to a two-dimensional time-scale joint representation that is highly redundant. The time-bandwidth product of the CWT is the square of the input signal and for most applications, which seek a signal description with as few components as possible, this is not desirable. To overcome this problem *discrete wavelets* have been introduced. Discrete wavelets are not continuously scalable and translatable but can only be scaled and translated in discrete steps. This is achieved by modifying the wavelet representation in equation (A.3) to produce [87]:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{a_0^j}}\phi\left(\frac{t - kb_0a_0^j}{a_0^j}\right) \tag{A.15}$$

Although it is called a discrete wavelet, it is normally a piecewise continuous function. In equation (A.15), j and k are integers and $a_0(>1)$ is a fixed dilation step. The translation factor b_0 depends on the dilation step. The effect of discretizing the wavelet is that the time-scale space is now sampled at discrete intervals. A typical value is a_0=2 so that the sampling of the frequency axis corresponds to *dyadic sampling*. This choice of value is also very suited for use in binary computers which speeds up processing times. For the translation factor a value of b_0 is chosen so that dyadic sampling is on the time axis as well. The concept of dyadic sampling is shown in figure A.3. When discrete wavelets are used to transform a continuous signal the result is a series of wavelet coefficients which are referred to as the *wavelet series decomposition*. An important issue in the decomposition is the subsequent reconstruction of the signal. In [87] it is shown that a necessary and sufficient condition for the stable reconstruction is that the energy of the wavelet coefficients must lie between two possible bounds:

$$A\|f\|^2 \;\leq\; \sum_{j,k} |\langle f, \psi_{j,k}\rangle|^2 \;\leq\; B\|f\|^2 \tag{A.16}$$

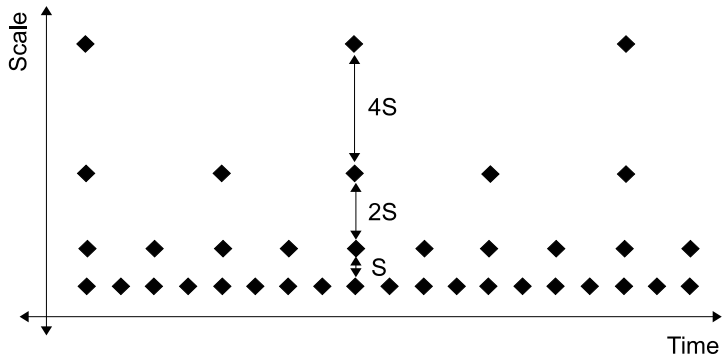


Figure A.3: Dyadic Grid

where $\|f\|^2$ is the energy of *f(t)*, $A > 0$, $B < \infty$ and *A,B* are independent of *f(t)*. When equation (A.16) is satisfied, the family of basis functions $\phi_{j,k}$ with j, $k \in \mathcal{Z}$ is referred as a *frame* with frame bounds A and B. When $A = B$ the frame is *tight* and the discrete wavelets behave exactly like an orthonormal basis. When $A \neq B$ exact reconstruction is still possible at the expense of using a *dual frame*. In a dual frame discrete wavelet transform, the decomposition wavelet is different from the reconstruction wavelet.

The next step in removing the redundancy from the wavelet transform is to ensure that the discrete wavelets are orthonormal. This can only be done with discrete wavelets which can be made orthogonal to their own dilations and translations. This is achieved by a special choice of mother wavelet as shown in equation (A.17):

$$\psi_{j,k}(x)\psi^*_{m,n}(x)dt = \begin{cases} 1 & \text{if } j = m \text{ and } k = n \\ 0 & otherwise \end{cases} \tag{A.17}$$

An arbitrary signal can be reconstructed by summing the orthogonal wavelet basis functions, weighted by the wavelet transform coefficients [86].

Orthogonality is not essential in the representation of signals. The wavelets need not be orthogonal and in some applications the redundancy can help to reduce the sensitivity to noise or improve the shift invariance of the transform [60]. This is a disadvantage of discrete wavelets: the resulting wavelet transform is no longer shift invariant, which

means that the wavelet transforms of a signal and of a time-shifted version of the same signal are not simply shifted versions of each other.

Dilations and translations of the "Mother function," or "analyzing wavelet" Φ(x) define an orthogonal basis for the wavelets:

$$\Phi_{s,l}(x) = 2^{\frac{-s}{2}}\Phi(2^{-s}x - l) \tag{A.18}$$

The variables s and l are integers that scale and dilate the mother function Φ(x) to generate wavelets, such as a Daubechies wavelet family. The scale index s indicates the wavelet's width, and the location index l gives its position. Notice that the mother functions are rescaled, or "dilated" by powers of two, and translated by integers. What makes wavelet bases especially interesting is the self-similarity caused by the scales and dilations. Once we know about the mother functions, we know everything about the basis.

To span the data domain at different resolutions, the mother wavelet is used in a scaling equation:

$$W(x) = \sum_{k=0}^{N-2}(-1)^k c_{k+1}\Phi(2x + k) \tag{A.19}$$

where *W(x)* is the scaling function for the mother function Φ(x), and c_k are the wavelet coefficients. The wavelet coefficients must satisfy linear and quadratic constraints of the form:

$$\sum_{k=0}^{N-1} C_k, \qquad \sum_{k=0}^{N-1} C_k C_{k+2l} = 2\delta_{l,0} \tag{A.20}$$

where δ is the delta function and l is the location index. This means that the above sum is zero for all l not equal to zero, and that the sum of squares of all coefficients is two [88].

One of the most useful features of wavelets is the ease with which one can choose the defining coefficients for a given wavelet system. In Daubechies' original paper [115], specific families of wavelets were developed that were very good at representing polynomial attributes in systems. The Haar wavelet, shown in figure A.4 is an even simpler

type of wavelet and it is often used for educational purposes.

The Haar wavelet is an example of a mother wavelet ϕ and is shown mathematically

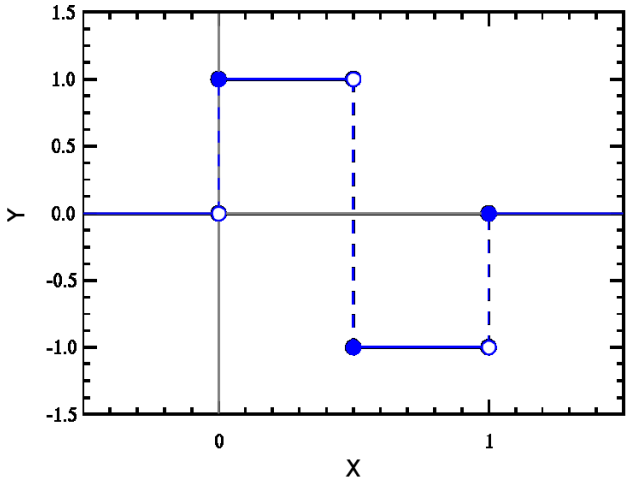


Figure A.4: Haar Wavelet

in equation A.21:

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ -1 & \text{if } \frac{1}{2} \leq x \leq 1 \\ 0 & otherwise \end{cases} \tag{A.21}$$

and the scaling function (father wavelet) is given by:

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & otherwise \end{cases} \tag{A.22}$$

Therefore $\psi(x) = \phi(2x - 1)$. The functions:

$$\psi_{m,n}(x) = 2^{\frac{m}{2}} \psi\left(2^m x - n\right), \qquad m, n \in \mathcal{Z} \tag{A.23}$$

form a basis in the Hilbert space $L_2(\mathcal{R})$. This means that every function $f \in L_2(\mathcal{R})$ can be expanded with respect to this basis. If m is restricted to m = 0,1,2,. . .and:

$$n = 0, 1, 2, \ldots, 2^m - 1 \tag{A.24}$$

the Hilbert space $L_2[0,1]$ is obtained.

This class of wavelet function is constrained, by definition, to be zero outside of a small interval. This is what makes the wavelet transform able to operate on a finite set of data. This property is known as *compact support*. The recursion relation ensures that a scaling function ϕ is non-differentiable everywhere. Of course, this is not valid for Haar wavelets. The following table lists coefficients for two wavelet transforms, the Haar wavelet and the Daubechies-4 wavelet. It is helpful to think of the coefficients

Wavelet	$\mathbf{c}_0$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$
Haar	1.0	1.0	0.0	0.0
Daubechies-4	$\frac{1}{4}(1+\sqrt{3})$	$\frac{1}{4}(3+\sqrt{3})$	$\frac{1}{4}(3-\sqrt{3})$	$\frac{1}{4}(1-\sqrt{3})$

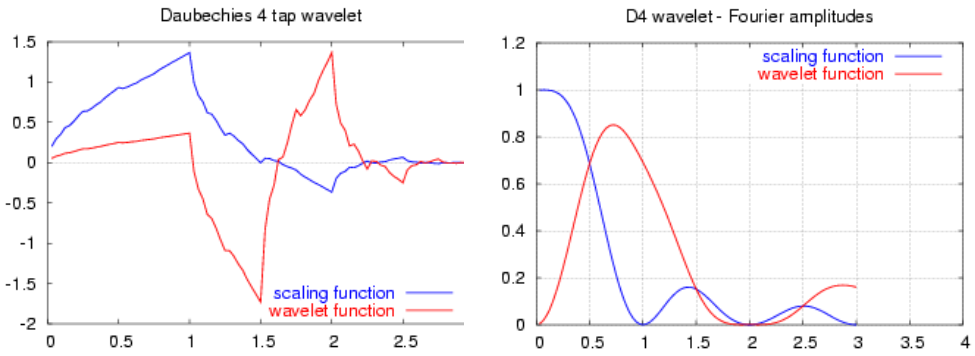
Table A.1: Coefficients for Two Wavelet Functions

$\{c_0,...,c_n\}$ as a filter. The filter or coefficients are placed in a transformation matrix, which is applied to a raw data vector. The coefficients are ordered using two dominant patterns, one that works as a smoothing filter (like a moving average), and one pattern that works to bring out the data's "detail" information. These two orderings of the coefficients are called a quadrature mirror filter pair in signal processing parlance. A more detailed description of the transformation matrix can be found in the literature [90].

A.2.1 Daubechie Wavelet

As previously outlined, the Daubechie wavelet is a compactly supported orthogonal system that is characterized by a maximal number of vanishing moments. In general the Daubechies wavelet is chosen to have the highest number A of vanishing moments (see section A.1.3) for a given support width $N = 2A$. Among the 2^{A-1} possible solutions the one is chosen whose scaling filter has extremal phase. The wavelet transform is also easy to put into practice using the fast wavelet transform. The Daubechie wavelet is used in this research due to its suitability for representing discontinuities and sharp peaks present in power system signals [91].

The Daubechie wavelets are not defined in terms of the resulting scaling and wavelet functions; in fact, they are not possible to write down in closed form. The graphs below are generated using the cascade algorithm, a numeric technique consisting of simply inverse-transforming [1 0 0 0 0 ...] an appropriate number of times. The spectra of the first four even taps are shown in figure A.5 with both the wavelet and scaling functions as well as the fourier transform amplitudes of the functions.



(a) Scaling and Wavelet Functions

(b) Amplitude of Frequency Spectra

Figure A.5: Scaling and Wavelet functions as well as Amplitudes of the Frequency Spectra of the Daubechie-4 Tap Wavelet

A.2.2 Band-Pass Filtering

With the redundancy removed, the number of wavelets required by the transform to represent the signal is still too large. Even with discrete wavelets, an infinite number of scalings and translations are required to calculate the transform. In order to make the wavelet transform practical, the signal has to be represented by a finite number of wavelets.

The translations of the mother wavelet are limited by the duration of the signal under investigation which provides an upper bound. The remaining variable in the analysis is the dilation: how many scales are required to analyze the given signal. In section A.1.1 we discussed the admissibility condition that stated that the Fourier transform of the

mother wavelet ψ(x) vanishes at zero frequency:

$$|\psi(\omega)|^2|_{\omega=0} = 0 \tag{A.25}$$

This means that the mother wavelet must have a band-pass like spectrum. This observation is crucial as we know from Fourier theory [92] that a compression in time is equivalent to stretching the frequency spectrum and shifting it upward:

$$Ff(at) = \frac{1}{|a|}F\left(\frac{\omega}{a}\right) \tag{A.26}$$

Therefore, a time compression of a wavelet by a factor of 2 will stretch the frequency spectrum of the wavelet by a factor of 2 and will also shift the frequency components up by a factor of 2. With this property, the finite spectrum of a signal can be covered with the spectra of dilated wavelets in the same way a time-domain signal is covered with translated wavelets. To get good coverage of the signal spectrum the stretched wavelet spectra should touch each other, as if they were starting hand in hand as shown in figure A.6. This can be arranged by carefully specifying the mother wavelet as described in section A.1.1. In summary, if one wavelet can be seen as a band-pass

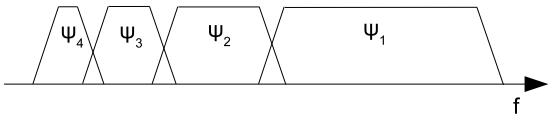


Figure A.6: Touching wavelet spectra

filter, then a series of dilated wavelets can be seen as a band-pass filter bank. If we look at the ratio between the center frequency of a wavelet spectrum and the width of this spectrum we see that it is the same for all wavelets. This ratio is normally referred to as the fidelity factor Q of a filter and in the case of wavelets, this refers to a $constant - Q$ filter bank.

A.2.3 Low Frequency Plugging

As discussed previously, when the time-domain representation of a signal is increased it is possible to only cover a corresponding proportion of the frequency spectrum. This

means that an infinite number of wavelets are required to cover the entire frequency spectrum.

The solution to this problem is not to attempt to cover the lowest frequencies but to use a low pass filter to essentially "plug" the frequency gap. This low pass filter action is determined by the *scaling function* which was discussed briefly in section A.1.2. The scaling function was introduced by Mallat [93]. Due to the low-pass nature of the scaling function spectrum it is sometimes referred to as an *averaging filter*.

If we look at the scaling function as being just a signal with a low-pass spectrum, then we can decompose it in wavelet components and express it as:

$$\varphi(x) = \sum_{j,k} \gamma(j,k)\psi_{j,k}(x) \tag{A.27}$$

Since the scaling function φ(x) was selected in such a way that its spectrum neatly

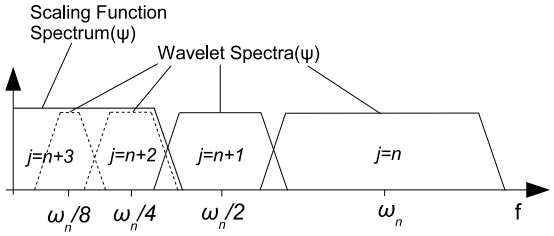


Figure A.7: Infinite Set of wavelets replaced by a single scaling function spectra

fitted in the space left open by the wavelets, the expression (A.27) uses an infinite number of wavelets up to a certain scale j (see figure A.7). This means that if the signal is analyzed using the combination of scaling function and wavelets, the scaling function itself takes care of the spectrum otherwise covered by all the wavelets up to scale j, while the rest is represented by the wavelets. In this way the number of wavelets required to represent the signal is now finite.

By introducing the scaling function we have circumvented the problem of the infinite number of wavelets and set a lower bound for the wavelet spectra. Of course when a scaling function is used instead of wavelets signal information is lost. That is to say, from a signal representation point of view no information is lost, since it will still be

possible to reconstruct the original signal, but from a wavelet-analysis point of view it is possible that valuable scale information may be discarded. The width of the scaling function spectrum is therefore an important parameter in the wavelet transform design. The shorter its spectrum the more wavelet coefficients will be generated and thus more scale information will be available. However, there will be practical limitations on the number of wavelet coefficients that can be handled. In the following sections, we will see that by using the discrete wavelet transform, this problem is more or less solved automatically.

A.2.4 Multiresolution Analysis

If we regard the wavelet transform as a filter bank, then the wavelet transform of a signal can be obtained by passing it through this filterbank. The outputs of the different filter stages are the wavelet transform and scaling function coefficients. Analyzing a signal by passing it through a filter bank is not a new idea and has been around for many years and is referred to as *subband coding*. It is widely used in computer vision applications and signal processing. The filter bank needed for subband coding can be

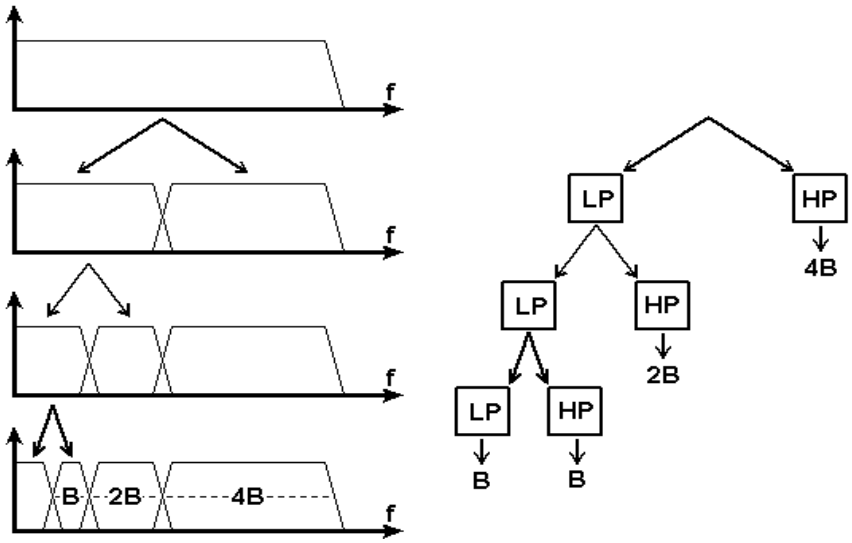


Figure A.8: Splitting the signal spectrum

built in several ways. One way is to build many band-pass filters to split the spectrum into frequency bands. The advantage here is that the width of every band can be

chosen freely and in such a way that the spectrum of the signal to analyze is covered at the points were it may be interesting. The disadvantage is that every filter has to be designed separately and this can be a time consuming process. Another way is to split the signal spectrum in two (equal) parts, a low-pass and a high-pass part. The high-pass part contains the smallest details (highest resolution) of the signal which are of most interest.

After this initial splitting two frequency bands are present. However, the low-pass part still contains some details and therefore this can be split again to extract the higher frequency detail. This process can be repeated until we are satisfied with the number of bands that have been created. In this way, an iterated filter bank has been created. Usually the number of bands is limited by the amount of data or computation power available. The process of splitting the spectrum is graphically displayed in figure A.8. The advantage of this scheme is that only two filter have to be designed, the disadvantage is that the signal spectrum coverage is fixed.

Looking at figure A.8 we can see what is left after the repeated spectrum splitting. What remains is a series of band-pass bands with doubling bandwidth and one low-pass band. Although in theory the first split gave a high-pass band and a low-pass band, in reality the high-pass band is a band-pass band due to the limited bandwidth of the signal. In other words, the same subband coding can be performed by feeding the signal into a bank of band-pass filters of which each filter has a bandwidth twice as wide as its left hand neighbor (the frequency axis runs to the right here) and a low-pass filter.

At the beginning of this section we stated that this is the same as applying a wavelet transform to the signal. The wavelet transform produces the band-pass bands with doubling bandwidth and the scaling function provides the low-pass band. From this we can conclude that a wavelet transform is the same thing as a subband coding scheme using a constant-Q filter bank [93]. In general, this kind of analysis will be referred to as a multiresolution analysis.

A.2.5 Pyramid Algorithm

It is of vital importance to choose the correct algorithm for the deconstruction and reconstruction of a given signal. There are a whole host of algorithms that may be used to adequately represent the signal. However, since typical power signals can be full of spikes and sharp discontinuities the best algorithm to use in this case is called the *pyramid algorithm*.

The pyramid algorithm operates on a finite set of N input datapoints x_0,x_1,. . .,x_{N-1}, where N is a power of two; this value will be referred to as the input block size. This data is passed through two *convolution functions*, each of which creates an output stream that is half the length of the original input. These convolution functions are filters, and one half of the output is produced by the low pass filter:

$$a_i = \frac{1}{2}\sum_{j=0}^{N-1} c_{2i-j+1}x_j \qquad i = 0, 1,. . .,\tfrac{N}{2}\text{-}1 \tag{A.28}$$

and the other half is produced by the high pass filter function:

$$b_i = \frac{1}{2}\sum_{j=0}^{N-1} (-1)^j c_{j-2i}x_j \qquad i = 0, 1,. . .,\tfrac{N}{2}\text{-}1 \tag{A.29}$$

where N is the input block size, c_j are the coefficients:

$$\mathrm{x} = (x_0, x_1, \ldots, x_{N-1}) \tag{A.30}$$

is the input sequence, and:

$$\mathrm{a} = (a_0, a_1, \ldots, a_{N/2-1}), \qquad \mathrm{b} = (b_0, b_1, \ldots, b_{N/2-1}) \tag{A.31}$$

are the output sequences. In the case of the lattice filter, the low and high pass outputs are usually referred to as the odd and even outputs, respectively. In many situations, the odd or low-pass outputs contains most of the information regarding the original input signal. The even, or high-pass outputs contains the difference between the true input and the value of the reconstructed input if it were to be reconstructed from only the information given in the odd output. In general, higher order wavelets (i.e. those

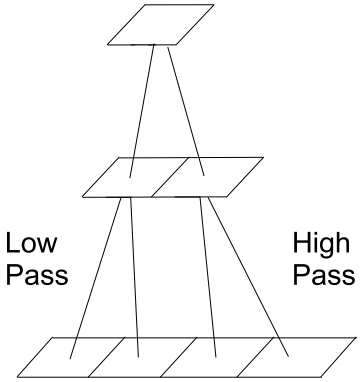


Figure A.9: Pyramidial structure showing low-pass and high-pass outputs derived from the algorithm which are concatenated to form a single vector

with more non-zero coefficients) tend to put more information into the odd output, and less into the even output. If the average amplitude of the even output is low enough, then half of the signal may be discarded without greatly affecting the quality of the reconstructed signal. An important step in wavelet based data-compression is finding wavelet functions that cause the even terms to be nearly zero. However, details can only be neglected for very smooth time series and smooth wavelet filters, a situation which is not satisfied for chaotic time signals.

To complete the discussion of the DWT, it is useful to look at how the wavelet coefficient matrix is applied to the data vector. The matrix is applied in a hierarchical algorithm, sometimes called a pyramidal algorithm. The wavelet coefficients are arranged so that odd rows contain an ordering of wavelet coefficients that act as the smoothing filter, and the even rows contain an ordering of wavelet coefficient with different signs that act to bring out the data's detail. The matrix is first applied to the original, full-length vector.

Then the vector is smoothed and decimated by half and the matrix is applied again. Then the smoothed, halved vector is smoothed, and halved again, and the matrix applied once more. This process continues until a trivial number of "smooth-smooth-smooth..." data remain. That is, each matrix application brings out a higher resolution of the data while at the same time smoothing the remaining data. The output of the DWT consists of the remaining "smooth" components, and all of the accumulated

"detail" components.

Figure A.9 shows a simplified pyramid structure which represents the two data vectors, smoothed and detailed, that are derived from each signal and subsequent sub-signals.

A.2.6 Implementation of Discrete Wavelet Transform

In equation (A.27) it was stated that the scaling function could be expressed in wavelets from minus infinity up to a certain scale j. If we add a wavelet spectrum to the scaling function spectrum we will get a new scaling function, with a spectrum twice as wide as the first. The effect of this addition is that the first scaling function can be expressed in terms of the second, because all the information required to do this is contained in the second scaling function. This can be expressed formally in the so-called multiresolution formulation [60] or two-scale relation [86]:

$$\varphi(2^j x) = \sum_k h_{j+1}(k)\varphi(2^{j+1}t - k) \tag{A.32}$$

The two-scale relation states that the scaling function at a certain scale can be expressed in terms of translated scaling functions at the next smaller scale. The first scaling function replaced a set of wavelets and therefore the wavelets in this set can also be expressed in terms of translated scaling functions at the next scale. More specifically, it may be written for a wavelet at level j:

$$\psi(2^j x) = \sum_k h_{j+1}(k)\varphi(2^{j+1}t - k) \tag{A.33}$$

which is the two-scale relation between the scaling function and the wavelet.

Since our signal $f(t)$" could be expressed in terms of dilated and translated wavelets up to a scale j-1, this leads to the result that $f(t)$ can also be expressed in terms of dilated and translated scaling functions at a scale j:

$$f(x) = \sum_k \lambda_j(k)\varphi(2^j t - k) \tag{A.34}$$

To be consistent with the notation we should, in this case, speak of discrete scaling functions since only discrete dilations and translations are allowed. If in this equation

the scale is stepped up to j-1, wavelets have to be added in order to keep the same level of detail. The signal $f(t)$ can then be expressed as

$$f(x) = \sum_k \lambda_{j-1}(k)\varphi(2^{j-1}t - k) + \sum_k \gamma_{j-1}(k)\psi(2^{j-1}t - k) \tag{A.35}$$

If the scaling function $\varphi_{j,k}$ and the wavelets $\psi_{j,k}$(x) are orthonormal or a tight frame, then the coefficients λ_{j-1}(k) and γ_{j-1}(k) are found by taking the inner products:

$$\lambda_{j-1}(k) = \langle f(x), \varphi_{j,k}(t) \rangle \tag{A.36}$$

$$\gamma_{j-1}(k) = \langle f(x), \psi_{j,k}(t) \rangle \tag{A.37}$$

If $\varphi_{j,k}$(t) and $\psi_{j,k}$(t) are replaced in the inner products by suitably scaled and translated versions of (A.32) and (A.33) and is subject to minor manipulation, keeping in mind that the inner product can also be written as an integration, we arrive at the important result [60]:

$$\lambda_{j-1}(k) = \sum_m h(m - 2k)\lambda_j(m) \tag{A.38}$$

$$\gamma_{j-1}(k) = \sum_m g(m - 2k)\gamma_j(m) \tag{A.39}$$

These two equations state that the wavelet and scaling function coefficients on a certain scale can be found by calculating a weighted sum of the scaling function coefficients from the previous scale. Now recall from the section on the scaling function that the scaling function coefficients came from a low-pass filter and recall from the section on subband coding how a filter bank was iterated by repeatedly splitting the low-pass spectrum into a low-pass and a high-pass part. The filter bank iteration started with the signal spectrum, so if we imagine that the signal spectrum is the output of a low-pass filter at the previous (imaginary) scale, then we can regard our sampled signal as the scaling function coefficients from the previous (imaginary) scale. In other words, our sampled signal $f(k)$ is simply equal to λ(k) at the largest scale.

In signal processing theory, a discrete weighted sum like the ones in (A.38) and (A.39) is the same as a digital filter and since the coefficients λ_j(k) are known to come from

the low-pass part of the split signal spectrum, the weighting factors $h(k)$ in (A.38) must form a low-pass filter. And since the coefficients γ_j(k) are known to come from the high-pass part of the split signal spectrum, the weighting factors $g(k)$ in (A.39) must form a high-pass filter. This means that (A.38) and (A.39) together form one stage of an iterated digital filter bank and from now on the coefficients $h(k)$ will be referred to as the *scaling filter* and the coefficients g(k) as the *wavelet filter* [94].

It has been shown that implementing the wavelet transform as an iterated digital filter bank is possible and from this point on we can speak of the *discrete wavelet transform* or *DWT*. Due to this, a useful bonus property of (A.38) and (A.39) can be exploited and is called the subsampling property. It can be seen from the equations that the scaling and wavelet filters have a step-size of 2 in the variable k. The effect of this is that only every other λ_j(k) is used in the convolution, with the result that the output data rate is equal to the input data rate. Although this is not a new idea, it has always been exploited in subband coding schemes and is useful as the output signals require no padding.

The subsampling property also solves the problem of how to choose the width of the scaling function spectrum. Everytime the filter bank is iterated, the number of samples for the next stage is halved so that in the end we are left with just one sample (in the extreme case). It will be clear that this is where the iteration definitely has to stop and this determines the width of the spectrum of the scaling function. Normally the iteration will stop at the point where the number of samples has become smaller than the length of the scaling filter or the wavelet filter (whichever is the longest), so the length of the longest filter determines the width of the spectrum of the scaling function. Figure A.10 illustrates the signal splitting procedure to produce the scaling and wavelet coefficients.

The scale of the signal, which is concerned with the relative proportions of the numerous frequency components in the signal, is determined by downsampling operations. The output of the cascaded filterbanks associated with the selected mother wavelet

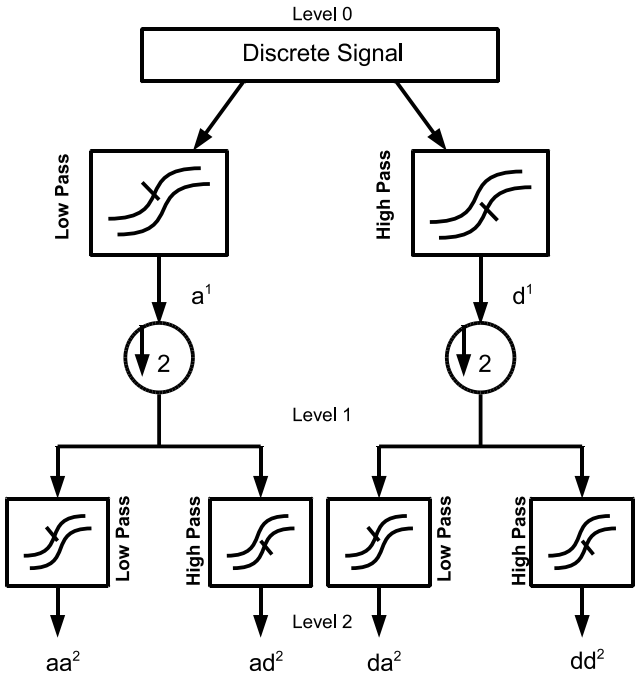


Figure A.10: Two Level Decomposition of the Discrete Wavelet Transform

and resolution levels produce details and approximations of the original signal. These approximations are then related to high frequency and low frequency components respectively. Computing the full wavelet packet decomposition (WPD) of a discrete time signal involves applying both wavelet (high-pass) and scaling (low-pass) filters to the discrete-time signal [x_1,x_2,x_3,...,x_n] and then recursively to each intermediate signal thereafter.

The procedure is illustrated in figure A.10 up to the second level of resolution.
The first level of decomposition of the discrete signal $x[n]$ of length N generates two frequency sub-bands. These are the approximation and detail coefficients, ${a_n}^1$=[${a_0}^1{a_1}^1{a_2}^1$,...,${a_{N-1}}^1$] and ${d_n}^1$=[${d_0}^1{d_1}^1{d_2}^1$,...,${d_{N-1}}^1$], respectively.

In essence, the a_n vector represents a smoothed version of the original signal with the d_n vector representing the more detailed (higher frequency) data that has been removed. The detailed coefficient output from the filters are known as $wavelet\ coefficients$ and determine the frequency content of a signal at a given time at a given resolution.

The second level of decomposition generates four frequency sub-bands using the same set of filters as the first resolution level. These four sub-bands are $aa^2[N/2]$ (low frequency approximations), $ad^2[N/2]$ (low frequency details), $da^2[N/2]$ (high frequency approximations), and $dd^2[N/2]$ (high frequency details). The second level low and high frequency details and approximations can be mathematically expressed as:

$$aa_n^2 = \sum_{k=0}^{\frac{N}{2}-1} g[k]a_n^1[2n-k] \tag{A.40}$$

$$ad_n^2 = \sum_{k=0}^{\frac{N}{2}-1} h[k]a_n^1[2n-k] \tag{A.41}$$

$$da_n^2 = \sum_{k=0}^{\frac{N}{2}-1} g[k]d_n^1[2n-k] \tag{A.42}$$

$$dd_n^2 = \sum_{k=0}^{\frac{N}{2}-1} h[k]d_n^1[2n-k] \tag{A.43}$$

where $g[k]$ and $h[k]$ are digital filter coefficients.

A.3 Wavelets in the Power System

Wavelets enjoy various properties including orthogonality, localization and fast computational algorithms. The time-frequency aspect of wavelets that can deconstruct a signal of interest into a function over the time-frequency plane that tells us *when* each *frequency* occurs is particularly useful. It is this ability to simultaneously locate frequency and time content of a signal that makes it suitable for oscillation source location.

This time-frequency attribute allows wavelets to be used for analyzing non-stationary time series and signals (e.g. speech signals). Another application of this time-frequency localization is that wavelets can describe the local features (such as steps) of a function and provide tools for studying change-points in statistics and edge detection in image processing. These change-points are of course analogous to event detection in power system dynamics.

A.3.1 Event Detection

Most of the events in a power system are routine in nature. Occasionally some abnormal events do occur which can cause significant damage to the system. The detection and identification of such dynamic events can provide an early alert, which can be used in conjunction with timely stabilizing controls to reduce cascading failures.

Dynamic events are critical to system reliability. They often cause significant frequency deviation and persist for a relatively long period of time [95] [96]. Their frequency content usually concentrates in a certain range. The range of electromechanical oscillations is typically 0.2Hz-0.8Hz and how to detect them is still challenging due to their dynamic nature.

With the use of PMU's and WAMS, it has been possible to better understand dynamics as well as review and coordinate relay settings and guide operating decisions. Studies in the literature [6] [97] describe methods based on digital filters to identify the onset of a dynamic event, referred to as a *triggering problem*.

In [98], the author examines the feasibility of using wavelet analysis on frequency signals to detect and analyze dynamic events in a power system. Wavelet analysis can yield information about the time of occurrence and the frequency content of individual events. Potential applications of wavelet detection include turning event recording on and off and triggering stability controls to control poorly damped modes. The estimated post-event damping is used to determine the end of dynamic events and to separate overlapping events. Some of the work later on will show how an event is triggered using wavelet analysis.

Additionally, wavelet analysis may be used to track certain modes as they evolve under varying system conditions. Implementation of this mode tracking system is essential for a successful solution to the source location problem as any mode analyzed must be consistent across a range of system conditions.

A.3.2 Frequency Characteristics of System

In Chapter 2 we discussed how a power system state, such as active power, can be considered to be the combination of a large number of aperiodic and oscillatory modes at any given time t. These modes are derived from the eigenvalues λ of the state matrix A and it is the *complex eigenvalues* that occur in conjugate pairs that describe oscillatory modes.

Instability occurs due to inadequate damping of any of the oscillatory (or aperiodic) modes that are present in a power system. The root of these unstable modes lie with the active components of the system i.e the generators. By measuring the active signals at the generators or on the transmission lines it is possible to extract information about the dominant system modes. Since it is the generators that are critical in mode behaviour, as they supply the oscillation energy, then it is wise to study generator behaviour during periods of instability.

Moreover, since there may be only a single or small number of troublesome modes in the system, then it makes sense to study generator outputs and line flows in terms of mode frequency i.e. focus on the mode frequency range of the generator output. This would be extremely useful as system modes are predominantly excited by corresponding generator outputs of similar mode frequencies. By comparing the mode anywhere in the system with the bandpassed generator outputs it is possible to get a clearer picture of cause and effect without the distraction of other unrelated frequencies.
Figure A.11 shows how a signal is broken down by the discrete wavelet transform. In this instance an N=4 digital filter is used which is composed of eight coefficients. The effect of the filtering process is to continually halve the signal bandwidth to allow analysis at smaller and smaller resolutions. The minimum resolution in this case is $16\times f_b$, where f_b is the reciporical of the total sampling length.

For example, figure A.11 refers to a data length of 32 points. If these were sampled every 0.5 seconds then the total sampling length is $0.5\times32 = 16$ seconds. Therefore,

the lowest frequency component in this is signal is $1/16 = 0.0625$Hz which is denoted as f_b. The highest resolution level achievable with 32 data points is resolution level 4 and corresponds to $16 \times f_b = 1$Hz with resolution level 3 corresponding to 0.5Hz an so on. Initially, the more familiar Fourier transform sounds like a promising candidate

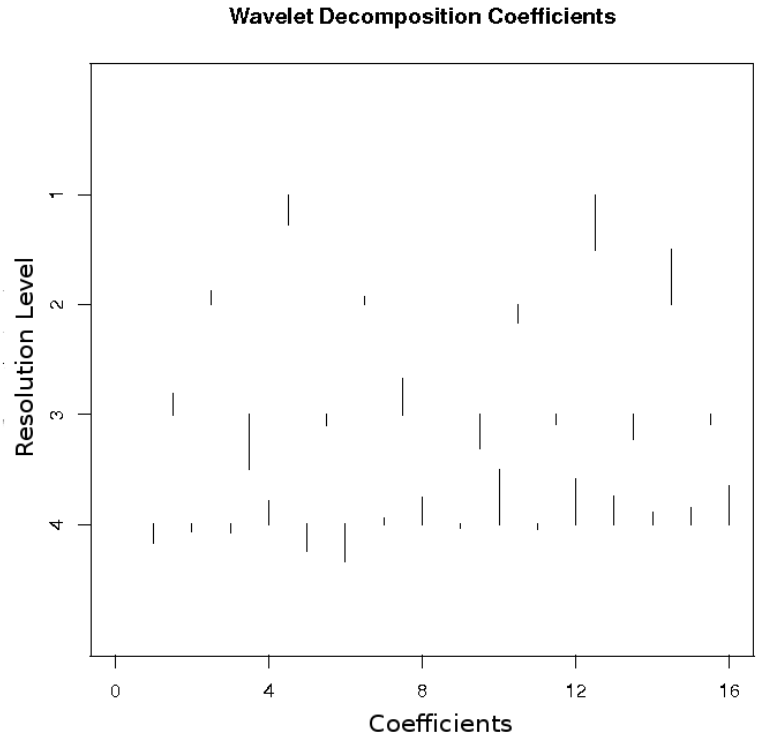


Figure A.11: Frequency bands and Wavelet Coefficients of model power signal with N=4 resolution

for a frequency representation of a power system signal however, this does not supply any temporal information i.e. exhibiting which frequencies are present *when*. In sectionr̃efsf2 we showed how the wavelets can be stacked and their coefficients displayed graphically to see which frequency bands are present throughout the signal. This is the method used from here on in with signal analysis. By deconstructing the active signals it is then possible to apply various statistical techniques to these transformed predictors and a dependent variable. The dependent variable in question could be a wavelet transformed frequency, active or reactive power signal at a specified location in the system or, as is used in this study, the decay time constants of the event modes in question.

A.3.3 Wavelets in Statistics

The properties of wavelets may be used to develop wavelet based methodologies for solving statistical problems. By representing a signal in terms of wavelets and their coefficients it is possible to extract more detailed information for use in statistical models that would otherwise be impossible. The primary focus of this study, is the incorporation of wavelets into regression models with the aim of producing better fitted statistical models in comparison to using raw data.

Wavelets may also be used in non-parametric regression to better represent the response variable in question. Depending on the statistical attributes of the variables, the wavelet transform may be used in both parametric and non-parametric regressions as the removal of redundant frequencies may allow the model to highlight important relationships that hitherto would have been undetected. Wavelets also benefit from their suitability to non-stationary stochastic processes as well as change-points in both direct and indirect data.

An interesting aspect of the wavelet transform is the ability of thresholding wavelet coefficients. Thresholding techniques may be used to extract relatively large coefficients that represent large changes in the signal at a given frequency. This would essentially be an analysis of changes in the system and may prove helpful in detecting significant predictors as well as in event detection as an event is usually characterized by sudden changes in dynamics [99].

Nonparametric Regression

Consider the nonparametric regression model:

$$y_i = f(x_i) + \epsilon_i, \qquad i = 1, \ldots, n = 2^J \tag{A.44}$$

where $x_i = i, n$, ϵ_i are independent and identically-distributed normal random errors (i.i.d), and $f(x)$ is a function. The problem is the estimation of $f(x)$ based on data y_i. Because wavelets provide sparse representations for a wide class of functions Donohoe and Johnstone [100] proposed wavelet shrinkage to take advantage of wavelets sparse

representations and efficiently estimate regression functions. Wavelet shrinkage works as follows. First compute the discrete wavelet coefficients of data y_1,...,y_n. Secondly, shrink the wavelet coefficients using the shrinking conditions in equations (A.46) (A.47) (A.48) and thirdly, construct an estimator of $f(x)$ by using the shrunk wavelet coefficients. Specifically, let $(y_{j,k})$, $(\theta_{j,k}$ and $(\epsilon_{j,k})$ be discrete wavelet coefficients of (y_i), $(f(x_i))$ and (ϵ_i), respectively. From equation (A.44) we yield:

$$y_{j,k} = \theta_{j,k} + \epsilon_{j,k} \qquad j = 0, 1, \ldots, J-1, \quad k = 0, 1, \ldots, 2^j - 1 \tag{A.45}$$

where $\epsilon_{j,k}$ are i.i.d normal, due to the orthogonality of discrete wavelet transformations.

Due to the wavelets sparse representations, there are a relatively small number of large $|\theta_{j,k}|$ and a large number of small $|\theta_{j,k}|$. Shrinking rules are used to select only those $y_{j,k}$ whose corresponding $\theta_{j,k}$ are of large magnitude. We use the selected $y_{j,k}$ to recover the large $\theta_{j,k}$ and reconstruct the function from the recovered wavelet coefficients as an estimator of f.

The two shrinking rules are the hard and soft shrinking rules which are shown below respectively:

$$\hat{\theta}_{j,k} = \sigma_h(y_{j,k}, \lambda) = y_{j,k} 1_{\{|y_{j,k}| > \lambda\}} \tag{A.46}$$

$$\hat{\theta}_{j,k} = \sigma_h(y_{j,k}, \lambda) = sign(y_{j,k})(|y_{j,k}| - \lambda)_+ \tag{A.47}$$

where λ is the threshold. The wavelet estimator $\hat{f}$ is the function constructed using $\hat{\theta_{j,k}}$ as wavelet coefficients.
Various ways are proposed to select the threshold λ. Universal threshold is defined by:

$$\lambda = \sigma\sqrt{2log(n)} \tag{A.48}$$

which is used to shrink $y_{j,k}$ at all levels. It is derived from the fact that with probability tending to one, the maximum $|\epsilon_{j,k}|$ is bounded by $\sigma\sqrt{2log(n)}$.

There are a host of other thresholding techniques including the SURE, block, FDR and cross-validation methods that are detailed in [101] [102] [103].

Appendix B

Logic Regression

B.1 Logic Regression Background

B.1.1 Boolean Logic Terminology

Equation (B.1) shows a typical Boolean statement that can be used to represent a value, state, condition etc. Some of the basic concepts of Boolean Logic will be clarified here with reference to equation (B.1).

$$L = \left[(X_1 \wedge X_2) \vee (\bar{X_3} \wedge X_4)\right] \wedge (X_5 \vee \bar{X_6} \vee X_7) \tag{B.1}$$

Values

Each predictor and response variable is designated either 1 or 0 (on or off, True or False etc) and is determined by setting a threshold.

Variables

Symbols denoted X_1,X_2,X_3 and Y_1,Y_2 etc. are variables that represent one of two possible values as explained above.

Operators

Operators are used to combine the variables into Boolean statements. The three different types of operators are AND $\wedge$, OR $\vee$ and NOT $^-$.

Expressions

The combination of values and variables with operators results in Boolean expressions. For example:

$$X_1 \wedge \bar{X}_2 \tag{B.2}$$

is a Boolean logic expression built from two variables and two operators. X_1 and $\bar{X}_2$ are called the operands of $\wedge$.

B.1.2 Rules and Laws of Boolean Algebra

Boolean algebra is an abstract algebraic discipline, axiomized by a set of equations that determine its properties. These are the rules and laws that apply to Boolean algebra alone and these are given below [104].

Boolean Algebra Precedence Algebra

As in standard algebra there are precedence rules in place to determine which operations are carried out first to determine the value of an expression. The precedence rules are as follows:

1. Complement operators ($^-$) are applied first.
2. AND operators ($\wedge$) are then applied.
3. OR operators ($\vee$) are then applied.
4. Operators are applied in precedence from left to right.

Therefore, $X_1 \wedge \bar{X}_2$ means $X_1 \wedge (\bar{X}_2)$ and not $\overline{(X_1 \wedge X_2)}$.

Boolean Algebraic Laws

The following laws are stated without proofs but their validity can be determined by considering all the combinations of the dichotomous binary values.

- Associative Operators

$$(X_1 \wedge X_2) \wedge X_3 = X_1 \wedge (X_2 \wedge X_3) \tag{B.3}$$

$$(X_1 \vee X_2) \vee X_3 = X_1 \vee (X_2 \vee X_3) \tag{B.4}$$

- Commutative Operators

$$X_1 \wedge X_2 = X_2 \wedge X_1 \tag{B.5}$$

$$X_1 \vee X_2 = X_2 \vee X_1 \tag{B.6}$$

- Double Complement Law

$$Y = \overline{\bar{X}}; \qquad X = \bar{Y} \tag{B.7}$$

- Identity Laws

$$X \wedge 1 = X \tag{B.8}$$

$$X \vee 0 = X \tag{B.9}$$

- Null Laws

$$X \vee 0 = 0 \tag{B.10}$$

$$X \wedge 1 = 1 \tag{B.11}$$

- Complement Laws

$$X \wedge \bar{X} = 0 \tag{B.12}$$

$$X \vee \bar{X} = 1 \tag{B.13}$$

- Idempotent Laws

$$X \wedge X = X \tag{B.14}$$

$$X \vee X = X \tag{B.15}$$

- Distributive Laws

$$X_1 \vee (X_2 \wedge X_3) = (X_1 \vee X_2) \wedge (X_1 \vee X_3) \tag{B.16}$$

$$X_1 \wedge (X_2 \vee X_3) = (X_1 \wedge X_2) \vee (X_1 \wedge X_3) \tag{B.17}$$

- de Morgan's Laws

$$\overline{(X_1 \vee X_2)} = (\bar{X}_1 \wedge \bar{X}_2) \tag{B.18}$$

$$\overline{(X_1 \wedge X_2)} = (\bar{X}_1 \vee \bar{X}_2) \tag{B.19}$$

- Absorption Laws

$$X_1 \wedge (X_1 \vee X_2) = X_1 \tag{B.20}$$

$$X_1 \vee (X_1 \wedge X_2) = X_1 \tag{B.21}$$

B.1.3 Representations of Logic Statements

Disjunctive Normal Form

A widespread notation of logic statements in engineering circles is called *disjunctive normal form*, which is a special case of Boolean expression. This form is a Boolean expression expressed as $\vee$ combinations of $\wedge$ terms. For example, the logic terms in equation (B.22) are rewritten in equation (B.23) in disjunctive normal form.

$$(A \wedge \bar{B}) \wedge [(C \wedge D) \vee (E \wedge (\bar{C} \vee F))] \tag{B.22}$$

$$(A \wedge \bar{B} \wedge C \wedge D) \vee (A \wedge \bar{B} \wedge E \wedge C) \vee (A \wedge \bar{B} \wedge E \wedge F) \tag{B.23}$$

Equation (B.23) contains all the information in equation (B.22) with the only difference being that equation (B.22) is factored. The DNF, as we will see later, has certain advantages over the more compact form of equation (B.23).

Logic Trees

By using appropriate brackets, any Boolean expression can be shown figuratively as a logic tree. For example the logic equation (B.22) can be split into various tree branches and leaves as shown below:

$$\underbrace{\underbrace{(A \wedge \bar{B})}_{1} \wedge \underbrace{[\underbrace{(C \wedge D)}_{2} \vee \underbrace{(E \wedge \underbrace{(\bar{C} \vee F)}_{3})}_{4}]}_{5}}_{6} \tag{B.24}$$

The above expression can be understood as an "AND" statement generalized from the Boolean expressions $(A \wedge \bar{B})$ AND $(C \wedge D) \vee (E \wedge (\bar{C} \vee F))$. The latter part of this statement can be understood as an "OR" statement, generated from the Boolean expressions $C \wedge D$ OR $E \wedge (\bar{C} \vee F)$, and so on. This enables us to represent any logic term in a binary logic tree. The logic tree for equation (B.22) is shown in figure B.1. White letters on the black background denote a conjugate of the variable i.e. NOT that variable. The evaluation of the tree as a logic statement for a particular case occurs

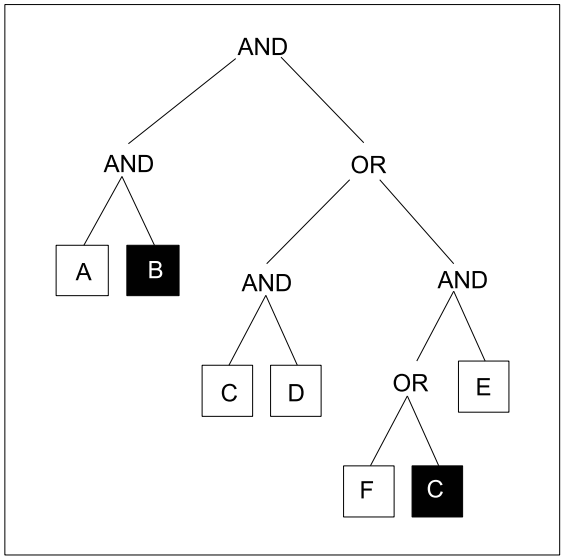


Figure B.1: Logic Tree representing the boolean expression from equation (B.22)

in a "bottom up" fashion. We use the following terminology and rules for Logic Trees similar to that used in a study of classification trees [105].

- The location for each element (letter, conjugate, letter or operator) in the tree is called a knot.
- Each knot either terminates in a zero or has two sub-knots.
- The two sub-knots are termed "children" with the initial knot being "termed" the parent.
- The knot that does not have a parent is called the root.
- The knots that do not have children are called the leaves.
- Leaves can only be occupied by letters or conjugate letters (predictors), all other knots are Boolean operators.

Since the representation of a Boolean expression as a logic term is not unique, neither is the logic tree representation. For example, the Boolean expression in equation (B.22) or (B.24) can also be written as:

$$((A \wedge \bar{B}) \wedge (C \wedge D)) \vee ((A \wedge \bar{B}) \wedge (E \wedge (\bar{C} \vee F))) \tag{B.25}$$

This statement leads to the tree shown in figure B.2. It is not just a matter of the complexity of the Boolean expression, since it can also be written as:

$$A \wedge \{\bar{B} \wedge [(C \wedge D) \vee (E \wedge (\bar{C} \wedge F))]\} \tag{B.26}$$

which would lead to yet another tree. A Boolean expression can be written in different

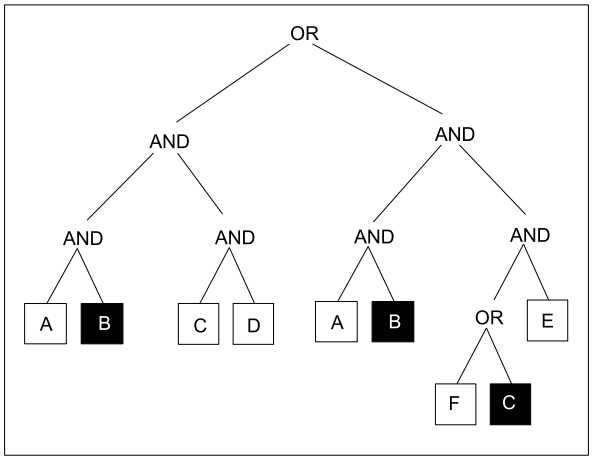


Figure B.2: Logic Tree representing the boolean expression from equation (B.25)

ways as a logic expression, but each logic expression corresponds exactly to one logic tree.

Generalized Logic Trees

At this point, it is possible to introduce a hierarchy of the leaves in a logic tree by assigning a number to each leaf (referred to as the depth of a leaf), and counting how often the types of operators change on the path between its parent and the root. For example, leaf "D" in the tree in figure B.2 has an "AND" as parent, which has an "OR" as parent etc. The entire sequence of links between leaf "D" and the root is $\wedge \longrightarrow \vee \longrightarrow \wedge$, which means there are three link changes (counting the parent of the leaf as a change number one), and hence leaf "D" has depth 3. The table below shows the depth for each leaf in the logic tree of figure B.1. Using the depth information form its

corresponding tree, it is possible to display every Boolean expression as a *generalized*

Leaf	A	$\bar{B}$	C	D	E	$\bar{C}$	F
Depth	1	1	3	3	3	4	4

Table B.1: Depth values for each Boolean variable in figure B.1

logic tree. On each "depth level" of a tree there is only one type of link (either $\wedge$ or $\vee$), and leaves of the same depth. Note that those trees are not binary in general and the generalized logic trees are neither sub nor superset of the logic trees as introduced in section B.1.3. Loosely speaking, a logic tree is to a Boolean expression what a generalized logic tree is to a Boolean expression with as many parenthesis removed as possible. Figure B.3 shows the generalized logic tree for the Boolean expression in

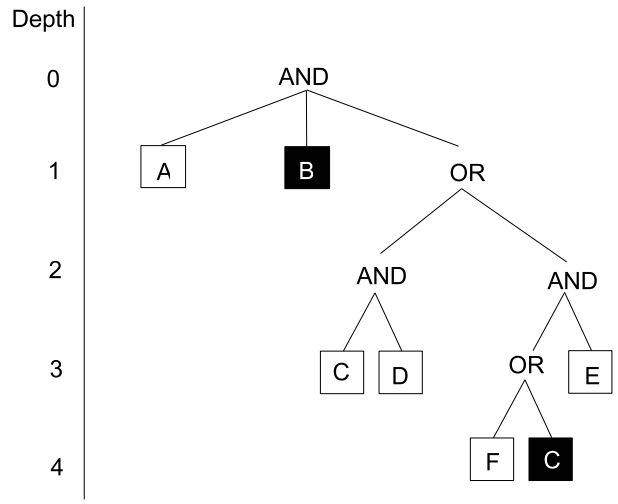


Figure B.3: Generalized logic tree representing equation (B.22) with depth levels

equation (B.22).

B.1.4 Equivalence of Logic Terms, Disjunctive Normal Forms, Logic Trees and Generalized Logic Trees

Initially, it is not clear whether or not the above stated representations of logic expressions are equivalent in the sense that the classes of logic expressions they represent are the same. We can show these classes are indeed the same, by establishing the following relations between classes:

$$Class(LTE) \subseteq Class(LTR) \subseteq Class(DNF) \subseteq Class(GLT) \tag{B.27}$$

where the above abbreviations are logic term LTE), logic tree (LTR), disjunctive normal form (DNF) and generalized logic tree (GLT).

Class(LTE) $\subseteq$ Class(LTR)

In the previous section we already outlined how to construct a logic tree from a given Boolean expression. The procedure is as follows:

1. Combine all pairs of letters in brackets to initial subtrees.
2. Combine single letters with subtrees in brackets to new subtrees.
3. Combine subtrees.

The steps of construction of the Logic Tree for the Logic Term in equation (B.22) are illustrated in figure B.4

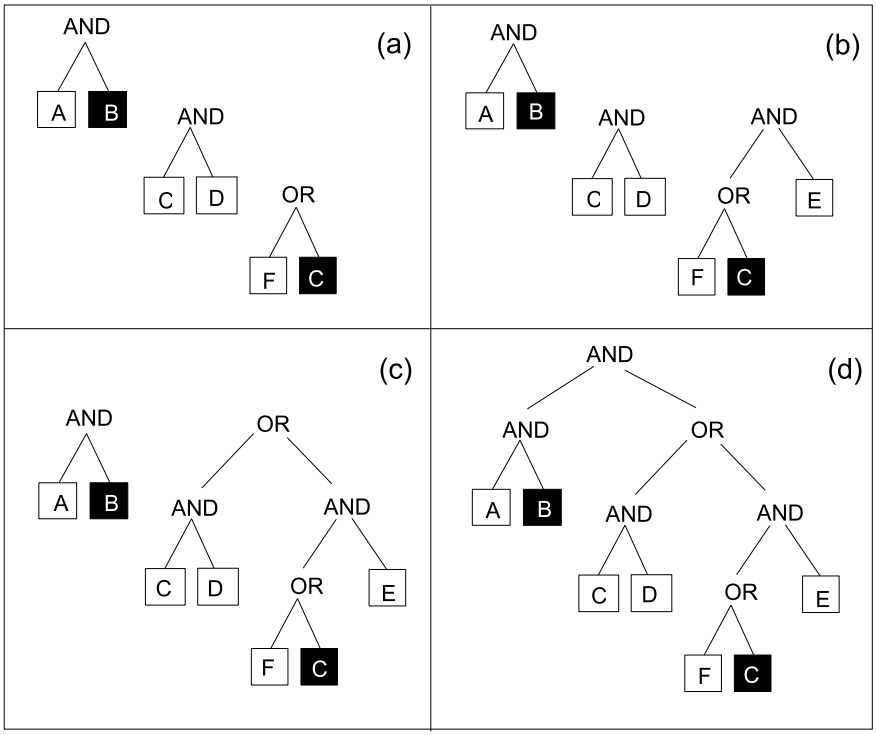


Figure B.4: Illustration of the construction of a logic tree in figure B.1 from the logic term equation (B.22)
:

Class(LTR) $\subseteq$ Class(DNF)

In this section we establish that the class of logic expressions represented by logic trees is a subset of the class of logic expressions represented by disjunctive normal forms. We show this by constructing an algorithm, that generates the disjunctive normal form of a Boolean expression from its logic tree. The algorithm requires that the knots are numbered as shown in figure B.3, starting at 1 with the root and with the left most knot on each level of the tree being numbered as a power of two.

A logic expression L in disjunctive normal form is a $\vee$-combination of $\wedge$-terms. This means that L is true if one of the $\wedge$-terms is true, which is the case if within the $\wedge$-term all predictors are true. Consider a logic tree with leaves $\{X_1, \ldots, X_k\}$ from "left to right" in the tree, independent of the level and not in the sequence.
The idea of the algorithm is to obtain the disjunctive normal form of the Boolean expression from this logic tree. The procedure is to split the set of leaves into subsets until we have sets of leaves that represent the $\wedge$-terms in the disjunctive normal form. We sequentially check the logic tree for $\vee$-operators, starting with the root. There are two subtrees of the root, with the left and right child of the root as their respective subtrees. The subtrees have the leaves $X_1, \ldots, X_j$ and $X_{j+1}, \ldots, X_k$ respectively. If the root is a $\vee$-operator, it is sufficient and and necessary for the entire tree to be true that at least if one subtree is true.

If the root is a $\wedge$ operator, both subtrees have to be true. In the former case we will consider the set of leaves $\{X_1, \ldots, X_j\}$ and $\{X_{1j+1}, \ldots, X_k\}$ independently, and check what kind of operators knot 2 and 3 are. In the latter case, check knot 2: if it is a $\wedge$, then both subtrees of knot 2 and the subtree with knot 3 as a root have to be true for the entire tree to be true.
If knot 2 is a $\vee$, then it is necessary for at least one subtree of the knot to be true in addition to the subtree with knot 3 as its root. Hence, if knot 2 is a $\vee$, we have to consider the sets $\{X_1, \ldots, X_l, X_{j+1}, \ldots, X_k\}$ and $\{X_{l+1}, \ldots, X_j, X_{j+1}, \ldots, X_k\}$. In other words, at every knot that is a $\vee$ operator we split the appropriate set of leaves,

until we have considered the last $\vee$. Then, if for any of those subsets all leaves are true, the entire tree is true.

Class(DNF) $\subseteq$ Class(GLT)

It is straightforward to construct a generalized Logic Tree from a Boolean expression in Disjunctive normal form. The root is a $\vee$ and its children are $\wedge$'s. The number of children the root has is simply the number of $\wedge$-terms in the Disjunctive normal form. All children of the $\wedge$'s in the tree are leaves, representing the predictors in the respective terms in the Disjunctive normal form.

Class(GLT) $\subseteq$ Class(LTE)

Since any generalized logic tree represents some logic term, there is nothing to show in this class relation.

B.1.5 Relationship between Logic forms and Decision Trees

At first glance logic trees seem to be quite similar to classification trees as introduced by Brieman et al [105]. However, there are some key differences between those types of trees.

In every classification tree, a leaf can be reached by a path through the tree, making decisions at every knot. If the tree is binary, these decisions reduce to checking whether or not the condition investigated at a particular knot is true or false. To reach a certain leaf, all conditions C_1,C_2,...,C_k along the path have to be satisfied (i.e. $C_1 \wedge C_2 \wedge, \ldots, \wedge C_k$ has to be true).

In general, there are multiple paths that reach a leaf that predict class 1. Since there are only two outcome classes (0 and 1), the collection of all paths P_1,P_2,...,P_l that reach a leaf predicting 1 is a complete description of the binary classification tree. Therefore, the tree predicts class 1 for a case if the Boolean equation:

$$L = P_1 \vee P_2 \vee \ldots \vee P_l \tag{B.28}$$

is true, where:

$$P_i = C_1^i \wedge C_2^i \wedge \ldots \wedge C_{k^i}^i \tag{B.29}$$

Hence every binary classification tree can be written as a Boolean equation in disjunctive normal form.

B.2 Search Algorithms

In the previous section it was shown that given a fixed number of predictors, there are only finitely many Boolean expressions that can yield different predictions. If we have k predictors, we showed that there are 2^k different prediction scenarios. However, given the values of the predictors, we would like to know how many logic trees there are that yield different predictions. If we have l cases i.e. a sufficient number of predictors, there might be up to 2^l different logic trees.

For example, if we have 100 cases and 10 predictors this could produce more than 10^{300} different logic trees. In addition to the fact we have to deal with such an array of logic trees, there seems to be no straightforward way to list all the logic trees that yield different predictions. In order to extract the good logic trees from the endless possibilities we have to use a search algorithm, the study of which is the subject of this section.

B.3 Moving in the search space

Within the search algorithm we define the neighbours of a certain logic tree to be the trees that can be reached from this logic tree by a single move. We allow the following moves:

- Alternating a leaf:

  We pick a leaf and replace it with another leaf at this position. For example, in figure B.5(b) the leaf B from the initial tree has been replaced with the leaf D^c. To avoid tautologies, if the sibling of a leaf is a leaf as well, the leaf cannot be replaced with its sibling or the complement of its sibling. It is clear that the

counter move to alternating a leaf is by changing the replaced leaf back to what it was before the move (i.e. alternating the leaf again).

- Changing $\wedge$'s and $\vee$'s

  Any $\wedge$ can be replaced with a $\vee$ and vice versa. For example, the operator at knot 1 in the initial tree in figure B.5(a) has been changed in figure B.5(e). These two moves complement each other as move and counter move.

- Branching (pruning)

  At any knot that is not a leaf, we allow a new branch to grow. This is done by declaring the subtree starting at this knot to be the right hand branch of the new subtree at this position and the left hand branch to be a leaf representing any predictor. These branches or sides are connected by a $\wedge$ or a $\vee$ at a location on the knot. For example, at knot 3 in the tree in figure B.5(e) a branch was grown (see figure B.5(f)). The counter-move to branching is called pruning. A leaf is trimmed from the existing tree. The subtree starting at the sibling of the trimmed leaf is "shifted" up to start at the parent of the trimmed leaf. This is illustrated in figure B.5(d).

- Splitting (deleting): Any leaf can be split by creating a sibling and determining a parent for those two leaves. For example, in figure B.5(c) the leaf C from the tree in figure B.5(b) has been split with leaf D^c as its new sibling. The counter move is to delete a leaf in a pair of siblings that are both leaves as illustrated in figure B.5(a).

In principle, a logic tree can be reached from any other logic tree in a finite number of moves even if one omits pruning and branching. In this sense, pruning and branching are not necessarily in the move set. However, their inclusion in the move set can enhance the performance of the algorithms that will now be introduced.

B.4 Greedy Search

Similar to the search algorithm in classification and regression trees [105], a greedy algorithm can be used to search for "good" logic trees. In the context of logic regression,

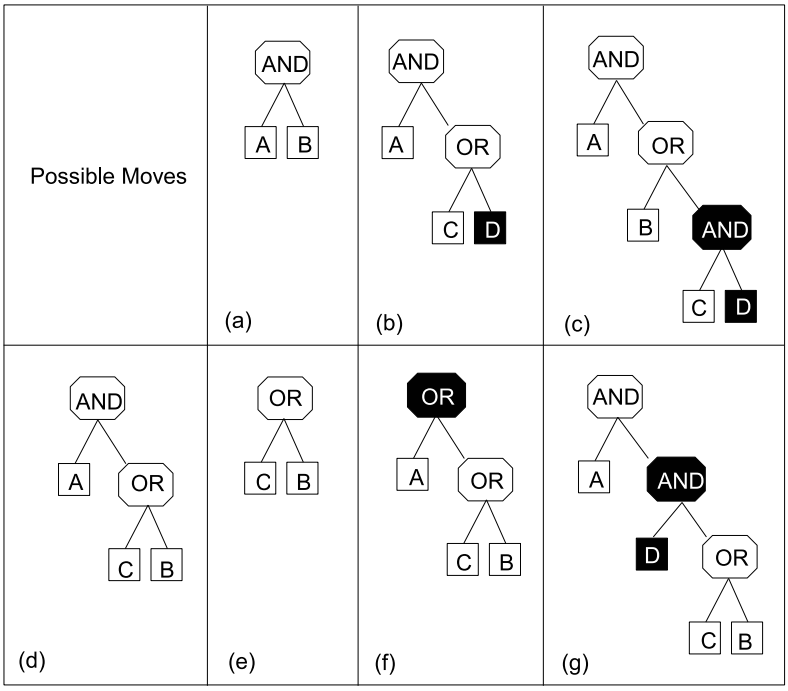


Figure B.5: Permissible Moves in the Tree Growing Process

the first step is simply to find a variable that, used as a single predictor, minimizes the scoring function. After this predictor is found, its neighbours (states that can be reached in a single move) are investigated. Subsequently, the new state is chosen as the state that:

1. Has a better score that the original state
2. Has the best score among the considered neighbours

If such a state does not exist the greedy algorithm stops, otherwise the neighbours of the new state are examined and the next state is chosen according to the above described criterion.

If, due to the stop criterion, the algorithm is unable to find a move that improves the score, there is no guarantee that the lowest possible state score will be reported. This can happen if the search gets stuck due to a better tree being attainable in two moves but not one. Another potential problem is that in the presence of noise the model may be over-fitted. This occurs when the true lowest score model is found but some untrue

neighbour trees seem to score better due to the presence of noise.
It is also noteworthy that in contrast to the greedy search for classification and regression trees, a greedy move for logic trees might actually result in a tree of lower or equal complexity, say for example by deleting a leaf or changing an operator respectively.

B.5 Simulated Annealing

A greedy search algorithm is very fast compared to a probabilistic search algorithm, such as the one that we will introduce in this section. However, the greedy search can lead to wrong results in certain cases as illustrated in [106]. The search can be trapped in a state that scores the best locally (i.e. all its neighbours have a worse score), but not globally. This unfortunate attribute can be avoided if the search algorithm is probabilistic.

There is an analogy between this type of probabilistic algorithm and phenomena found in the fields of statistical mechanics and condensed matter physics called quantum annealing. This is where the annealing term stems from [107].
Some more terminology is introduced in the next section along with some basic definitions regarding simulated annealing. The terminology used from here on is derived from [108]. The definitions and theorems were also taken from [108] as well as [109].

B.5.1 Terminology and Definitions

The annealing algorithm is defined on a state space S, which is a collection of individual states. Each of these states represents a configuration of the problem under investigation. The states are related by a neighbourhood system and the set of neighbour pairs in S defines a substructure $M = S \times S$.
The elements in M are called moves. Two states s and s' are called adjacent if they can be reached by a single move (i.e. $(s,s') \in M$). Similarly, $(s,s') \in M^k$ are said to be connected via a set of k moves. It is a requirement that the state space is finite and the size of the state space is fixed. However, it can be arbitrarily large, therefore this assumption does not result in loss of generality.

The following functions govern the search through the state space:

Score Function:

$$\in: S \to \mathcal{R}_+ \tag{B.30}$$

assigns a positive real number (score) to each state.

The score is understood as a measure of the quality of a state. In the following work, we always assume that lower scores are associated with states that represent better quality configurations. Since the state space is finite, there exists at least one with a minimal score. The score is denoted by $\in_0$.

Selection Probability Function:

$$\beta : S \times S \to [0, 1] \tag{B.31}$$

such that:

$$\forall_{(s,s')\in M}\beta(s, s^{'}) = 0 \tag{B.32}$$

$$\forall_{(s,s')\in M}\beta(s, s^{'}) \neq 0 \tag{B.33}$$

$$\forall_{s\in S} \sum_{s^{'}\in S} \beta(s, s^{'}) = 1 \tag{B.34}$$

Therefore, the selection probability is the probability that the state $s^{'}$ is proposed as the new state, given that the current state is s. The move set can now be defined as:

$$M = \{(s, s^{'}) \in S \times S : \beta(s, s^{'}) > 0\} \tag{B.35}$$

The move set M is symmetric if:

$$\forall_{s\in S}\forall_{s^{'}\in S}[(s, s^{'}) \in M \Rightarrow (s^{'}, s) \in M] \tag{B.36}$$

Acceptance Function: The acceptance function:

$$\alpha : \mathcal{R}_{+}^{3} \rightarrow (0,1) \tag{B.37}$$

assigns a positive probability to a pair of scores and a positive real number called the *temperature*.

The acceptance function decides whether or not the proposed state will be accepted as the new state. Note that for any fixed temperature, the probability only depends on the scores of the current and proposed state.

Transition Probability: The transition probability is a function:

$$\tau : \mathcal{R}_{+}^{3} \rightarrow [0,1] \tag{B.38}$$

which is defined as:

$$\tau(s, s^{'}, t) = \begin{cases} \alpha(\epsilon(s), \epsilon(s^{'}), t) \times \beta(s, s^{'}) & \text{if } j = m \text{ and } k = n \\ 1 - \sum_{s^{''} \in S} \alpha(\epsilon(s), \epsilon(s^{''}), t) \times \beta(s, s^{''}) & otherwise \end{cases}$$

Therefore, the transition probability $\tau(s, s^{'}, t)$ can be understood as the probability that the next step is a move to state $s^{'}$, given that the current state is s and the temperature is t. The probability that the state after n moves is $s^{'}$, given the current state s and the temperature t, is denoted by $\tau(s, s^{'}, t)$.

A process that possesses the above property is called a $MarkovProcess$. A sequence of events as a special case of such a Markov process is called a $Markovchain$. A Markov chain in which the transitional probabilities between the pairs of states are constant throughout the process is called *homogeneous*. A Markov chain is called *irreducible* if any state in the chain is connected to any other state by only a finite number of moves, i.e. if:

$$\bigcup M^{k} = S \times S \tag{B.39}$$

A Markov chain is called *aperiodic* if for every state s the greatest common divisor of all integers $n \geq 1$ with $\tau(s, s^{'}, t) > 0$ is equal to 1.

B.5.2 Properties of Markov Chains

Theorem 1 below is usually referred to as the Chain Limit Theorem. It states that an irreducible and aperiodic (homogeneous) Markov chain has a limiting distribution.

Theorem 1

For each irreducible and aperiodic chain there exists a density function:

$$\pi : S \times \mathcal{R}_{+} \rightarrow (0, 1) \tag{B.40}$$

in s for any given $t > 0$, with:

$$\pi(s, t) = \lim_{n \rightarrow \infty} \tau_n(s^{'}, s, t) \tag{B.41}$$

(independent of $s^{'}$) and satisfying the following equations:

$$\sum_{s^{'} \in S} \pi(s^{'}, t)\tau(s^{'}, s, t) = \pi(s, t) \tag{B.42}$$

$$\sum_{s \in S} \pi(s, t) = 1 \tag{B.43}$$

Hence if we constructed an irreducible and aperiodic (homogeneous) Markov chain for the annealing algorithm (i.e. run the chain at a fixed temperature), the distribution of states sampled approaches a limit. However, the search through the state space should yield low scoring states.

Theorem 2

An irreducible and aperiodic chain with a symmetric move set has the property:

$$\forall_{s \in S} \left[\epsilon(s) \neq \epsilon_0 \Rightarrow \lim_{t \downarrow 0} \pi(s, t) = 0 \right] \tag{B.44}$$

if it has an acceptance function α satisfying:

$$\epsilon \geq \epsilon^{'} \rightarrow \alpha(\epsilon, \epsilon^{'}, t) = 1 \tag{B.45}$$

$$\epsilon > \epsilon^{'} > \epsilon^{''} \rightarrow \alpha(\epsilon, \epsilon^{'}, t) \times \alpha(\epsilon^{'}, \epsilon^{''}, t) = \alpha(\epsilon, \epsilon^{''}, t) \tag{B.46}$$

$$\epsilon < \epsilon^{'} \rightarrow \lim_{t \downarrow 0} \alpha(\epsilon, \epsilon^{'}, t) = 0 \tag{B.47}$$

Hence, if the equations in (B.47) are satisfied, the likelihood of a non-optimal scoring state in the limiting distributions tends to zero as the temperature tends to zero. Therefore, if the annealing run is a sequence of homogeneous Markov chains with decreasing temperatures, the search is guided toward optimal scoring states.

The requirements set out in (B.47) only affect the acceptance function and do not impose any restraints on β or M. In general it is quite easy to construct a state space with a symmetric moveset that guarantees irreducibility and aperiodicity for the chain in the search algorithm. The desirable properties of the chains as stated in Theorem 2 can be achieved by choosing the right acceptance function.

Theorem 3

The only acceptance functions $\alpha(\epsilon, \epsilon^{'}, t)$ that are:

- differentiable in $\epsilon^{'}$
- whose values depend on t and the difference of ϵ and $\epsilon^{'}$
- that satisfy the conditions of Theorem 2

have the form:

$$\alpha(\epsilon, \epsilon^{'}, t) = min\{1, e^{-(\epsilon^{'} - \epsilon)c(t)}\} \tag{B.48}$$

where $c(t)$ is a negative, monotonic and continuous function satisfying:

$$\lim_{t \downarrow 0} c(t) = -\infty \tag{B.49}$$

In this work $c(t) = -1/t$ is used, yielding the acceptance function:

$$\alpha(\epsilon, \epsilon^{'}, t) = min\{1, e^{(\epsilon^{'} - \epsilon)/t}\} \tag{B.50}$$

This acceptance function has been used comprehensively in the literature [110]. This is presumably the case due to the fact that simulated annealing originated in condensed matter physics where the above acceptance function bears a striking resemblance to the Boltzmann distribution which characterizes a system of particles in thermal equilibrium.

B.5.3 Practical Aspects of Simulated Annealing

When the simulated annealing algorithm was implemented for the Logic Regression methodology, some practical aspects of running the search had to be considered.

- In theory, trees of any size can be grown but considering that the models need to be interpreted it makes sense to limit the model sizes. Therefore, the tree sizes have to be limited to a certain number of leaves. The maximum number of leaves for any given tree is 16.

- At the beginning of a simulated annealing run at high temperatures, virtually all moves are accepted. Towards the end almost every proposed move is rejected. Somewhere inbetween those points is the *crunchtime* which is where the search should spend most of its time looking for viable paths through the state space.

- To speed up the simulated annealing algorithm the number of moves that have been accepted when running a Markov chain at fixed temperature is recorded. If the number reaches a pre-determined threshold the simulation exits the Markov chain (even if the number of specified iterations has not been reached) and lowers to the next temperature. This avoids spending too much simulation time at the start searching through a plethora of random models.

  Additionally, the lowest temperature in the run is specified before the annealing algorithm is started. This ensures that the model continues searching in the state space long after all moves have been rejected. This criterion can also be achieved

by halting the search when no moves have been accepted in a substantial number of chains. Again, this avoids running the algorithm to its specified end even though no improvement can be gained.

- Every logic tree has a finite number of neighbours. Especially toward the end of a run at very low temperatures where very few moves get accepted. Since simulated annealing is a probabilistic search, a move might get rejected several times before it gets accepted. In order to avoid any computational bottlenecks, all states visited and their corresponding scores are tabulated. Therefore, in deciding whether or not to accept a certain move, the trees of the proposed model have to be evaluated only once, which speeds up the search, especially at low temperatures.

- When implementing a simulated annealing run a number of parameters have to be specified. One of the key parameters is the annealing temperature range which consists of a high (starting) temperature and a low (finishing) temperature. As well as this a cooling scheme is implemented which determines the total number of chains that are run at constant temperatures. The goal is to make sure individual Markov chains run close to their limiting distributions and that they cool sufficiently slowly over a reasonable number of iterations in a chain.

  The temperatures are usually lowered in equal increments on a log_{10} scale. The number of chains between two subsequent powers of 10 depend on the data being used but the number is usually between 25 and 100. The length of the individual chains is usually between 10,000 and 100,000 though the number of iterations may need to be increased for larger datasets.

B.6 Logic Models

Previously, search algorithms were introduced that were used to find low scoring logic trees from an array of states in a state space. In this section, logic models are introduced in terms of their role in the logic regression framework. A differentiation is made here between logic models with single trees and logic models with multiple trees. The models

involving one logic tree are for classification and regression problems. In this case, an attempt is made to find a Boolean expression L that classifies each response into two possible classes. In other words, a classification rule is sought.

$$C = I_{\{L \ is \ True\}} \tag{B.51}$$

In the latter case, a characteristic of each of the two subpopulations is modelled such as the mean of a variable within a subpopulation. In this case, a logic rule and its parameters are sought out to optimize the model such that:

$$\mu = a + b \times I_{\{L \ is \ True\}} \tag{B.52}$$

In both cases, a meaningful measure of model performance i.e. scoring function has to be defined which is largely dependent on the type of problem being examined. The one tree logic model can be generalized to include logic models with multiple trees as given by:

$$\mu = a + \sum_{i}^{k} b_i \times I_{\{L_i \ is \ True\}} \tag{B.53}$$

The multiple tree models are discussed later which shows how different trees are scored via their logic regression coefficient.

B.6.1 Classification Problems

A common statistical problem is to classify members of a population in one of two possible categories solely based upon a set of predictive variables measured for each member of the population. For example, it is believed that certain types of cancer are partly caused by genetic abnormalities. The aim is to find out if it is possible to predict which class (cancer/no cancer) a person belongs to given a list of genes that are expressed in that person. The type of model that could be fitted in this example might be, "if gene A and B are expressed", or "gene C but not D", then it is more likely that that person belongs to the class "cancer".

In classification problems, rules are used to assign each case into two possible classes

say 0 or 1. A Boolean expression L predicts a class C for each case via $C_{predicted} = I_{\{LisTrue\}}$. The scoring function in this case would simply be the total number of misclassifications resulting from this prediction n_{mc}, or the misclassification rate defined as the total number of misclassifications divided by the total number of cases ($\epsilon = n_{mc}/n$).

Weights are also commonly used in situations were one class is much more common than the alternate class or when the types of misclassification are not equally severe. An example of this would be predicting someone is healthy when they are sick as this has more profound consequences than if the reverse were predicted.

B.6.2 Regression Problems

In regression problems that involve a single tree, rules are sought that characterize parameters associated with measurements in two subpopulations. If the response is binomial, the parameter of interest can be the odds of belonging to class 0 versus class 1. If the response is continuous, the parameter can be the average response in each of the two classes. An example of this would be a car sales survey. A good model in terms of car sales could be the following: A person who works professionally and has a professional degree, or is retired and lives in Monaco, belongs to a population which spends approximately 40,000 on average on a new car. Otherwise, this person belongs to a population which spends on average 25,000 on a new car. The two most important regression models are linear regression and logistic regression which are covered in the next section.

Residual Sum of Squares

In the case of a continuous response, the means of the two subpopulations are modelled, an example being the average amount spent on a car. It is assumed that the true underlying model for the measurements is taken to be:

$$Y = \beta_0 + \beta_1 I_{\{L \ is \ True\}} + \epsilon \qquad with \quad \epsilon \sim N(0, \sigma^2) \tag{B.54}$$

where L is a Boolean expression that determines the means of the response variable Y in the two subpopulations: the mean E[y] of the response variable in class 0 (when L=0) is β_0, which is β_1 smaller than the mean of the response variable in class 1 (when L=1). For a given L, the model is fit using the method of least squares. The goal of the residual least squares is to find estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the expression $\sum_i \left(\beta_0 + \beta_1 I_{\{L_i=1\}} - Y_i\right)^2$. The scoring function used in this case is the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{B.55}$$

where $(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 I_{\{L \ is \ True\}}$ are fitted values of the model under consideration. In this case, the fitted values are simply $\hat{\beta}_0 = \hat{Y}_{\{L \ is \ False\}}$ and $\hat{\beta}_1 = \hat{Y}_{\{L \ is \ True\}} - \hat{\beta}_0$.

Logistic Log-Likelihood

In the case of a binary response a logistic regression can be used. Instead of predicting the outcome for each case as in the classification case, we can model the probability (log of the odds) of the response belonging to class 1 or class 0. The true underlying model is assumed to be:

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 I_{\{L \ is \ true\}} \tag{B.56}$$

where π is the probability for a certain case to be in class 1 and L is a Boolean expression that determines these probabilities. The log of the odds (define here or in logit bit) of being in class 1 when L=0 is β_0, which is β_1 smaller than the log odds of being in class 1 when L=1. As a scoring function, deviance is used which is defined to be twice the difference between the maximum achievable log likelihood and that attained under the fitted model [71]. (See section GLMs.)

In the case of a single logic term, and hence tree, the fitted probabilities of being in

class 1 given $L = 0$ and $L = 1$ respectively are:

$$\hat{\pi}_0 = \frac{n_{01}}{n_0} \tag{B.57}$$

$$\hat{\pi}_1 = \frac{n_{11}}{n_1} \tag{B.58}$$

where n_{ij} is the number of cases for which we predict i and the response is $j (i, j \in \{0, 1\})$. The deviance is given in [71] then simplifies to:

$$\begin{aligned} DEV &= 2 \times \left[n_{00} \left(\frac{n_0}{n_{00}} \right) + n_{01} \left(\frac{n_0}{n_{01}} \right) + \ldots \right. \\ &\quad \left. \ldots + \left(\frac{n_0}{n_{10}} \right) + n_{11} \left(\frac{n_0}{n_{11}} \right) \right] \end{aligned}$$

Although only linear and logistic regression has been discussed here, any regression model can be used as long as a meaningful scoring function can be devised. In particular, this could include the generalized linear model (GLM).

It is important to note that the single tree case can be extended to handle multiple regression problems. Again, this extension is only permissible if a suitable scoring function is used. Equations (B.60) show the extended regression equations for both linear and logistic regression:

$$Y = \beta_0 + \beta_1 I_{\{L_1 \ is \ true\}} + \ldots + \beta_p I_{\{L_p \ is \ true\}} + \epsilon \quad with \epsilon \sim N(0, \sigma^2) \tag{B.59}$$

$$log \left(\frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 I_{\{L_1 \ is \ true\}} + \ldots + \beta_p I_{\{L_p \ is \ true\}} \tag{B.60}$$

Again, both residual sum of squares and deviance are used to score the models respectively. Further extensions of the logic regression can be found in [111] and [112] the latter of which contains examples on polychotomous regression which may be of interest in this work in the future.

B.7 Model Fitting

B.7.1 Single Tree Model Fitting

As discussed in section B.2, there are two search algorithms used to find suitable models for the data. Both the simulated annealing and greedy search involve a stepwise search through the state space. At every step, the Boolean expression at that point is altered according to a permissbale set of rules detailed in section B.5.1. Given a new tree, we then calculate the new score and decide whether or not to accept the model based on the new score and old score. The temperature is also adjusted if a simulated annealing algorithm is used to allow the model time to search in optimum search space.

In some models, a re-fitting of the parameters may take place. In the case of the linear regression type model, a change in the Boolean predictor L will also change the regression parameters β_0 and β_1. These are then used to calculate the goodness of fit i.e. residual sum of squares.
In the models presented later, the Markov chain is started at a standard temperature normally around 2 to 3°C and terminates at 0.1 to 0.01°C. The temperature of the simulated annealing run is normally decremented in steps of $\frac{1}{25}$ on the log_{10} scale. Each Markov chain at a fixed temperature is normally fixed at between 50,000 and 100,000 iterations.

In general, the median score of the model drops as temperature is decreased, as well as the variability of the model scores. Inevitably, there is always noise present in the data and once the true model is reached i.e. the number of misclassifications is at a minimum, the simulation will overfit the data and start predicting a portion of that noise. This phenomenon manifests itself by allowing the search to visit tree's that produce a lower score than that of the true model. As a result, overfitted models capture all the true variables and interactions but will also have additional predictors and interactions that are not part of the true model. Methods for dealing with noisy data and overfitting are discussed later in this chapter.

B.7.2 Fitting Continuous Predictors to a Model

In certain cases, it is possible to incorporate a continuous variable in its raw format, into the model. For example, let the binary predictors be $X_{b,1}, \ldots, X_{b,k}$, and the continuous predictors be $X_{c,1}, \ldots, X_{c,k}$. If we consider a logistic regression model, the model could be fitted such that:

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 I_{\{L \ is \ true\}} + \beta_p I_{\{L_p \ is \ true\}} + \beta_{c,1} X_{c,1} + \ldots + \beta_{c,k} X_{c,k} \quad \text{(B.61)}$$

where $L_1, \ldots, L_p$ are logic trees generated from the binary predictors. The model is fit exactly as described in section (fitting models) by simply including the continuous predictors when calculating the score.

However, one of the main advantages of logic regression models with exclusively binary predictors is their simplicity. In this case, the response and predictors are easily interpretable as they belong to one class or the other. Even though including raw variables in the model may decrease the deviance it may not be a desirable thing to do if ease of interpretation is important.

B.7.3 Dichotomizing Continuous Predictors

By using a splitting rule, continuous predictors can be dichotomized before model fitting takes place. An example of this is the use of regression stumps which are the result of a single step fit of a regression tree. A continuous predictor is split at the value that minimizes a certain criterion, for example the binomial log-likelihood or the misclassification rate. After dichotomizing all continuous variables can then be included with all the other binary predictors in the model search.

B.8 Removing Redundancy from Logic Trees

As we have seen previously in section B.7.1, the resulting logic trees from a simulated annealing run can contain redundancy. In order to remove redundancy in logic trees a method to detect if a tree can be simplified or if two trees share the same underlying

Boolean expression would be useful. A number of methods exist ranging from truth tables to Boolean algebra manipulation.

B.8.1 Truth Tables

If we assume there are k predictors involved in a logic tree, then since each predictor is binary there are, at most, 2^k different cases to consider. The prediction of a logic tree can be checked using a truth table that presents the prediction of a tree for the 2^k different cases. This method allows us to determine whether two trees are equivalent but does not allow any simplification of the trees.

B.8.2 Simplification using Algebraic Laws

A tree can also be simplified if the algebraic laws from section B.1.2 are applied to the logic terms that represent a logic tree. An example of this is as follows. If we take a logic term such as:

$$L = [A \vee (C \wedge (B \wedge \hat{C}))] \vee [B \wedge (C \vee \hat{B})] \tag{B.62}$$

Using the algebraic laws, we get:

$$A \vee (C \wedge (B \wedge \bar{C})) \equiv A \vee ((B \wedge \bar{C}) \wedge C)) \qquad \textit{commutative operators} \tag{B.63}$$

$$\equiv A \vee (B \wedge (\bar{C} \wedge C)) \qquad \textit{associative operators} \tag{B.64}$$

$$\equiv A \vee (B \wedge (\bar{C} \wedge \bar{\bar{C}})) \qquad \textit{double complement} \tag{B.65}$$

$$\equiv A \vee (B \wedge 0) \qquad \textit{complement law} \tag{B.66}$$

$$\equiv A \qquad \textit{identity law} \tag{B.67}$$

By similar deduction $B \wedge (C \wedge \bar{B}) \equiv B \wedge C$. Hence $L = A \vee (B \wedge C)$, which is the underlying Boolean term.

B.8.3 Simplification using Algebraic Laws

B.8.4 Removing Redundancy via Numerical Means

Each model or state S has a score $\epsilon(S)$ with the lowest global score denoted by ϵ_0. If a tree in a model contains redundancy then this tree could be simplified without changing the score of the model. To find the simplest model among the best scoring models, one can distinguish all the best scoring models by giving each of those models a bonus score, relative to its complexity:

$$A \vee (C \wedge (B \wedge \bar{C})) \equiv A \vee ((B \wedge \bar{C}) \wedge C)) \qquad \textit{commutative operators} \tag{B.68}$$

$$\forall_S : [\epsilon(S) > \epsilon_0] \Rightarrow \epsilon_{new}(S) = \epsilon(S) \tag{B.69}$$

$$\forall_S : [\epsilon(S) > \epsilon_0] \Rightarrow \epsilon_{new}(S) = \epsilon_0 - bonus(S) \tag{B.70}$$

where bonus(S) is a positive function of the state S, that rewards states for low complexity. In the current logic regression algorithms the inverse of the number of leaves is used as a bonus score to rate the simplest models.

B.9 Statistical Inference and Model Selection

Using simulated annealing allows the model with the lowest score to be determined using a range of acceptance and score functions. However, in the presence of noise it is known that the model tends to overfit the data. In this section some methods are discussed that attempt to solve the overfitting problem. In the comparison of different models a method to model complexity needs to be devised. In this case, we use the total number of logic trees as a measure of model complexity and call in *model size*.

B.9.1 Cross-Validation

In a cross-validation (CV) step we want to assess how well the best model of size k performs in comparison to models of different sizes. In order to do this the dataset is split into m equally sized groups. For each of the m groups the procedure is as follows. Remove the cases from group i from the dataset. Find the best scoring model of size k, using only the data from the remaining $m - 1$ groups and score the cases in group

i under this model. This yields score ϵ_{ki}. The cross-validated score for model size k is $\epsilon_k = \frac{1}{m}\sum_i \epsilon_{ki}$.

An alternative method to CV is the training/test set approach. If sufficient data is available, cases can be randomly assigned to two groups with pre-determined sizes, using one part of the data as a training set and the other part as a test set. Instead of using the entire dataset in the model fitting and model evaluation process, models are fitted using the training set and the optimum model size is deduced by scoring those models using the independent test set.

B.9.2 Randomization

In the previous sections it was shown how to find the best scoring model of a certain class using a greedy search or a simulated annealing. The search will always find the best model but the question of whether or not the model has been overfitted still needs to be answered.

For any model class present in the logic regression methodology the best scoring model first has to be determined. The null hypothesis to test is then: "There is no signal in the data". If the null hypothesis were true, then the best model fit on the data with the response randomly permutated should yield the same score as the best model fit on the original data. This procedure can be repeated as often as desired with the proportion of scores better than the score of the best model (with original data) as the p-value. This is evidence against the null hypothesis.

B.9.3 Optimal Model Size

For any model class used in the logic regression methodology (linear regression, logistic regression etc.), it is essential that the best scoring model is found with score ϵ^*. The best scoring model for a range of sizes (0 to k) is also found to determine the optimum size for the model. Randomization tests are used to this end to determine the optimal model size to produce the best fitting model.

The null hypothesis in each test is: "The optimal model has size j, the lower score

obtained by models of larger sizes is due to noise". Here, j can be each model size we consider ($j \in \{0, \ldots, k\}$. Let us assume the null hypothesis is true and the optimal model size is j, with score ϵ_j. For a model with p trees there can be up to 2^p classes. If we imagine a model with two trees then this would correspond to four classes. We now randomly permute the response within each of those classes.

The model of size j is still considered the best with a score ϵ_j. If the best model (of any size) is now developed, it will have a score ϵ^{**} as least as good but usually better than ϵ_j. However, this score is in part due to the noise in the data which is causing the overfitting.
If the null hypothesis were true, and the model of size j was indeed the best, then ϵ^* would be a sample from the same distribution as ϵ^{**}. The distribution can be estimated as closely as desired by repeating this procedure multiple times. On the other hand, if the best model had a size larger than j, then the randomization would yield, on average, lower scores than the average obtained from randomizing the response of the model of optimal size (larger than j).

The procedure is summarized as follows. A sequence of randomization tests is carried out starting with a test using the null model. This is the test to determine the signal and noise contents of the data. The best model of size 1 is then conditioned (randomization) to generate randomization scores. Conditioning is then performed on the best model of size two, etc. By comparing the distributions of the randomization scores, an informed decision can be made on the best model size to use.

Bibliography

[1] McNabb. P, Wilson. D, Bialek. J, *"Dynamic Model Validation of the Icelandic Power System using WAMS based Measurements of Oscillatory Stability"* PSCC, Glasgow, 2008

[2] Wilson. D, Hay. K, McNabb. P, Trehern. J, *"Identifying Sources of Damping Issues in the Icelandic Power System"*. IEEE.InProceedings, PSCC, Glasgow, 2008.

[3] P. McNabb, N. Bochkina, D. Wilson, J. Bialek *"Oscillation Source Location using Discrete Wavelet Transforms"*. IEEE.InProceedings, Transmission and Distribution Conference, New Orleans, 2010.

[4] P. McNabb, N. Bochkina, D. Wilson, J. Bialek *"Oscillation Source Location using Logic Regression"*. IEEE.InProceedings, IEEE Smart Grids Europe, Gothenburg, 2010.